Article

Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy

https://doi.org/10.1038/s41586-022-04817-8

Received: 24 February 2021

Accepted: 27 April 2022

Published online: 22 June 2022

Check for updates

Mark Chaffin[1], Irinna Papangeli[2], Bridget Simonson[1], Amer-Denis Akkad[2], Matthew C. Hill[1,3], Alessandro Arduini[1], Stephen J. Fleming[1,4], Michelle Melanson[5], Sikander Hayat[2], Maria Kost-Alimova[5], Ondine Atwa[1], Jiangchuan Ye[1], Kenneth C. Bedi Jr[6], Matthias Nahrendorf[3,7], Virendar K. Kaushik[5], Christian M. Stegmann[2], Kenneth B. Margulies[6], Nathan R. Tucker[8] & Patrick T. Ellinor[1,3,9]✉

Heart failure encompasses a heterogeneous set of clinical features that converge on impaired cardiac contractile function[1,2] and presents a growing public health concern. Previous work has highlighted changes in both transcription and protein expression in failing hearts[3,4], but may overlook molecular changes in less prevalent cell types. Here we identify extensive molecular alterations in failing hearts at single-cell resolution by performing single-nucleus RNA sequencing of nearly 600,000 nuclei in left ventricle samples from 11 hearts with dilated cardiomyopathy and 15 hearts with hypertrophic cardiomyopathy as well as 16 non-failing hearts. The transcriptional profiles of dilated or hypertrophic cardiomyopathy hearts broadly converged at the tissue and cell-type level. Further, a subset of hearts from patients with cardiomyopathy harbour a unique population of activated fibroblasts that is almost entirely absent from non-failing samples. We performed a CRISPR-knockout screen in primary human cardiac fibroblasts to evaluate this fibrotic cell state transition; knockout of genes associated with fibroblast transition resulted in a reduction of myofibroblast cell-state transition upon TGFβ1 stimulation for a subset of genes. Our results provide insights into the transcriptional diversity of the human heart in health and disease as well as new potential therapeutic targets and biomarkers for heart failure.

Given the immense public health burden of heart failure, understanding the underlying mechanisms at the molecular level has been at the forefront of cardiovascular research. One common cause of heart failure, dilated cardiomyopathy (DCM), manifests as dilation of the left ventricle (LV) with systolic dysfunction but normal LV thickness[5]. Conversely, hypertrophic cardiomyopathy (HCM) is a heterogeneous disease characterized by a thickening of the LV wall and is often caused by genetic mutations in sarcomere genes[5]. Considering the complex aetiology of heart failure, molecular phenotyping combined with careful examination of clinical phenotypes may yield insights into disease progression. For example, tissue-level comparisons of RNA and protein have uncovered disease-specific transcriptional programmes[3,4,6]. Using recent advances in single-cell and single-nucleus RNA-sequencing technologies, we sought to understand these transcriptional changes at single-cell resolution[7–9].

Expression profiles of cardiomyopathies

We performed single-nucleus RNA sequencing (snRNA-seq) in replicate on LV samples from 44 individuals, including 12 with DCM, 16 with HCM and 16 with non-failing (NF) hearts. After strict quality control, we removed 8 samples, leaving at least 1 technical replicate from 11 DCM, 15 HCM and 16 NF LVs (Extended Data Fig. 1 and Supplementary Tables 1 and 2). All the patients with DCM and HCM showed advanced cardiomyopathy requiring transplantation. The LV ejection fraction (LVEF) was less than 20% for patients with DCM, whereas 7 patients with HCM had a LVEF of less than 50% and 8 patients with HCM had a LVEF of 50% or more. Next, we removed low-quality nuclei (Extended Data Fig. 2), leaving a total of 592,689 nuclei, which aggregated into 21 clusters on the basis of transcriptional similarity (Fig. 1a,b). We identified genes that were expressed selectively in each cluster and estimated enrichment in biological processes on the basis of gene ontology[10,11] for these gene sets (Extended Data Fig. 3 and Supplementary Tables 3 and 4). On the basis of these marker genes and biological processes, we assigned cell-type labels to each cluster. Consistent with previous work[7], the most abundant cell types were cardiomyocytes, fibroblasts, endothelial cells, mural cells, macrophages and lymphocytes (Fig. 1c). Compared with NF hearts, DCM and HCM hearts showed a statistically credible decrease in cardiomyocytes and increases in vascular smooth muscle cells (VSMCs) and activated fibroblasts. Additionally, DCM hearts had statistically credible increases in fibroblasts, macrophages, and lymphocytes compared with NF hearts, whereas HCM hearts had

[1]Precision Cardiology Laboratory and the Cardiovascular Disease Initiative, The Broad Institute, Cambridge, MA, USA. [2]Precision Cardiology Laboratory, Bayer US LLC, Cambridge, MA, USA. [3]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. [4]Data Sciences Platform, The Broad Institute, Cambridge, MA, USA. [5]Center for the Development of Therapeutics, The Broad Institute, Cambridge, MA, USA. [6]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Center for Systems Biology, Department of Radiology, Massachusetts General Hospital, Boston, MA, USA. [8]Masonic Medical Research Institute, Utica, NY, USA. [9]Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, MA, USA. ✉e-mail: ellinor@mgh.harvard.edu

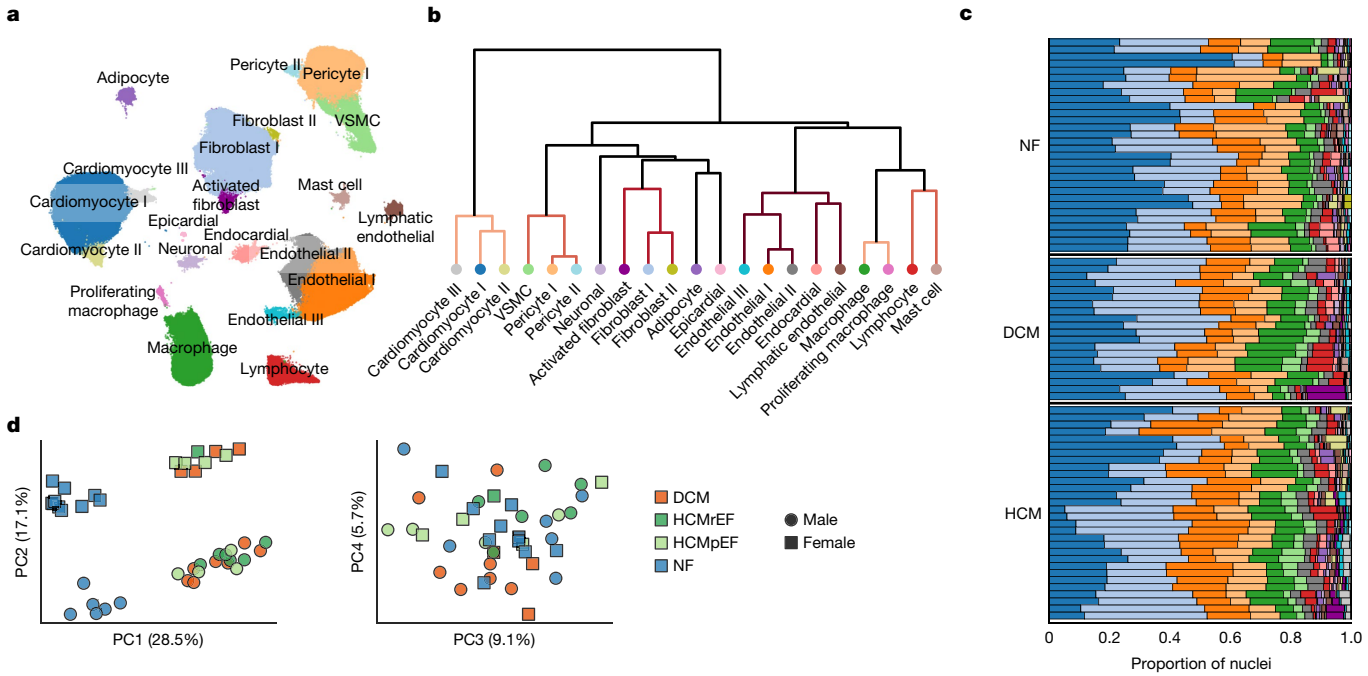


Fig. 1 | Cellular composition in the LV from healthy donors and those with cardiomyopathy. a, Uniform manifold approximation and projection (UMAP) representation of 592,689 nuclei isolated from LVs of 42 donors. **b**, Dendrogram demonstrating the similarity of cluster centroids. **c**, Stacked bar plot depicting the cell-type composition of each sample (n = 80; 1 or 2 technical replicates per patient) with colour coding reflecting cell types in **b**. **d**, PCA of pseudo-bulk snRNA-seq of LV samples from 42 donors by disease status and sex. The per cent variance captured by each principal component is shown in parentheses on each respective axis. HCMrEF, HCM with reduced LVEF; HCMpEF, HCM with preserved LVEF.

an increase in lymphatic endothelial cells compared with NF hearts (Extended Data Fig. 4 and Supplementary Table 5).

On the basis of principal component analysis (PCA) across combined data from all cell types per sample (pseudo-bulk sequencing), the largest source of transcriptional variation separated DCM and HCM hearts from NF hearts (Fig. 1d). The second largest source of variation divided donors by sex. This pattern was generally consistent when clustering patients within a given cell type (Extended Data Fig. 5). When performing PCA restricted to DCM and HCM hearts using pseudo-bulk data, we observed modest separation of patients based on LVEF across principal component 4 which explains around 7% of variation (Extended Data Fig. 5a).

We next sought to identify differentially expressed genes between cardiomyopathy and NF hearts across all cell types and within each cell type. We identified substantial changes in transcription when comparing DCM to NF and HCM to NF, but there were markedly fewer changes between DCM and HCM, consistent with the PCA results (Fig. 2a,b and Supplementary Tables 6, 7 and 8). Notably, the largest number of differentially expressed genes between each of the cardiomyopathy groups and NF hearts (false discovery rate (FDR) < 0.01) were found in fibroblasts, suggesting that the largest transcriptional differences occur in this cell type. There was a small set of genes that were differentially expressed between DCM and HCM hearts, primarily in cardiomyocytes, but these generally showed no patterns of enrichment in particular pathways. Overall, these results suggest a convergence to a common transcriptional profile in advanced cardiomyopathy requiring transplantation, consistent with previous observations based on proteomic analyses[4].

To further explore transcriptional differences amongst patients with cardiomyopathy, we compared HCM samples with preserved LVEF to HCM samples with reduced LVEF. We did not observe any differentially expressed genes in any cell-type at a FDR < 0.01. Additionally, in an exploratory analysis, we searched for genes with expression changes in cardiomyopathy that differed with sex. Despite limitations in the statistical power of this analysis, we identified a set of five genes that showed an interaction (FDR < 0.10) between cardiomyopathy status and patient sex, including the gene encoding thyroid hormone receptor β (*THRB*), suggesting that a small subset of genes may act differently in cardiomyopathy depending on sex (Fig. 2c).

On the basis of differential gene expression between cardiomyopathy and NF hearts, we performed pathway enrichment analyses by cell type to identify any systematic patterns in gene dysregulation (Fig. 2d and Supplementary Table 9). In general, pathway enrichments in DCM versus NF differential expression and HCM versus NF differential expression were similar. Of note, cardiomyocytes showed very little systematic up- or down-regulation in patients with cardiomyopathy. Conversely, we observed robust dysregulation in cardiomyopathy of several pathways in fibroblasts, including complement cascade, neutrophil degranulation, metabolism of vitamins and cofactors, potassium channels and neuronal systems. Similarly, multiple pathways in macrophages exhibited dysregulation among myopathic hearts, including neutrophil degranulation, generation of secondary messenger molecules and co-stimulation of the CD28 family.

Next, we sought to determine whether there were any distinct transcriptional differences associated with the genetic basis of DCM or HCM. We performed whole-genome sequencing on 40 of the 42 individuals included in this study to identify loss-of-function (LOF) mutations in known cardiomyopathy genes. We observed the expected enrichments of *TTN* LOF carriers in DCM (4 out of 10 patients), *MYBPC3* LOF carriers in HCM (3 out of 15 patients), and *MYH7* pathogenic variant carriers in HCM (5 out of 15 patients) (Extended Data Fig. 6a). However, there were no appreciable transcriptional differences in hearts from *TTN* LOF carriers compared with other patients with DCM, in hearts from *MYBPC3* LOF carriers compared with other patients with HCM, or in hearts from *MYH7* pathogenic variant carriers compared with other patients with HCM (Extended Data Fig. 6b–d).

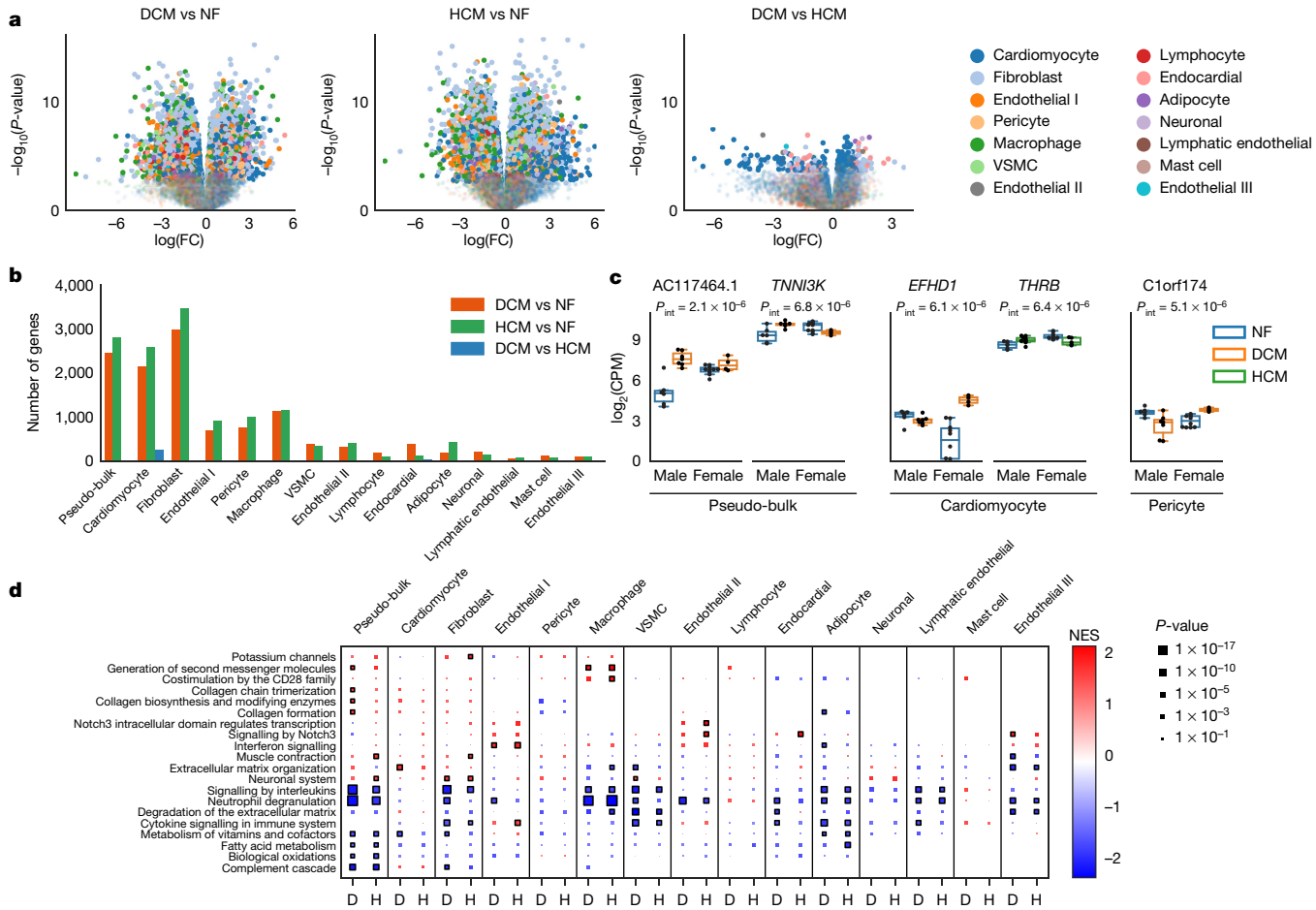


Fig. 2 | Transcriptional differences between NF and cardiomyopathy LVs. **a**, Log fold change and two-sided *P*-value for expression changes between DCM ($n = 11$) and NF ($n = 16$) (left), HCM ($n = 15$) and NF (centre), and HCM and DCM (right) hearts for each gene tested using limma–voom differential expression analysis (Methods). Genes are coloured by cell type with larger, opaque dots representing genes with FDR < 0.01 based on the Benjamini–Hochberg procedure. **b**, The number of significantly differentially expressed genes (FDR < 0.01) by cell type for each disease comparison in **a**. **c**, $\log_2$ counts per million (CPM) across patients for genes with a significant disease–sex interaction in either DCM ($n = 11$) versus NF ($n = 16$) or HCM ($n = 15$) versus NF ($n = 16$) hearts using limma–voom differential expression analysis (FDR < 0.1 based on the Benjamini–Hochberg procedure). In box plots, the center line respresents the median, box limits show upper and lower quartiles and whiskers span 1.5× the interquartile range. **d**, Reactome pathway enrichment for differential expression between DCM versus NF (D) and HCM versus NF (H) by cell type. The size of each square represents a two-sided *P*-value from gene set enrichment analysis (GSEA) and shading represents the normalized enrichment score (NES). Only pathways with a Benjamini–Hochberg FDR < 0.05 in both the GSEA and hypergeometric test for over-representation in at least one cell type are shown (Methods). Pathways with FDR < 0.05 in the GSEA test are denoted with a black outline. P_{int}, two-sided *P*-value for interaction between cardiomyopathy and sex.

Changes in cell subpopulations

To identify subpopulations within global cell types that may shift in composition in cardiomyopathy, we performed a formal sub-clustering analysis of the most abundant cell types. We observed at least one subpopulation with a statistically credible change in composition in either DCM or HCM within fibroblasts, endothelial cells, macrophages, VSMCs, and lymphocytes (Extended Data Figs. 7 and 8 and Supplementary Table 10). To characterize these subpopulations, we examined upregulated genes for each subpopulation compared with the remainder of nuclei from the given cell type (Extended Data Fig. 7 and Supplementary Table 11).

We observed a compositional increase in the endothelial subpopulation EC-*TMEM163* in both DCM and HCM compared with NF hearts (Extended Data Fig. 7c and Supplementary Table 10). This endothelial subpopulation displayed increased expression of previously reported angiogenic endothelial markers, including *KIT*, *NRP2*, *COL15A1*, *PCDH17* and *ITGA6*[12] (Supplementary Table 11). Similarly, we observed an increase in the lymphocyte subpopulation LC-*LINGO2* in DCM and HCM compared with NF hearts (Extended Data Fig. 7f and Supplementary Table 10). Compared with other lymphocytes, LC-*LINGO2* showed increased expression of several known natural killer cell markers including *KLRF1*, *GNLY*, *CD244* and *PRF1*[13] (Supplementary Table 11).

Among the macrophage populations, we identified four macrophage subpopulations, including a cluster of proliferating macrophages with a highly enriched cell cycle score and upregulation of many known cell cycle markers[14] (Extended Data Fig. 8a–d and Supplementary Table 11). Comparing the composition of macrophage subpopulations across patients, we observed a statistically credible reduction in the abundance of proliferating macrophages in patients with cardiomyopathy (Extended Data Fig. 8e and Supplementary Table 10). Immunofluorescence staining of the macrophage marker CD163 and the cell cycle marker MKI67 confirmed the enrichment of these proliferating macrophages in NF hearts (Extended Data Figs. 8f and 9a). By further interrogating the expression of prototypical marker genes of C-C chemokine receptor 2 (CCR2)-positive and CCR2-negative macrophages across these subpopulations, we observed that this population of proliferating macrophages more closely resembles reparative CCR2-negative resident

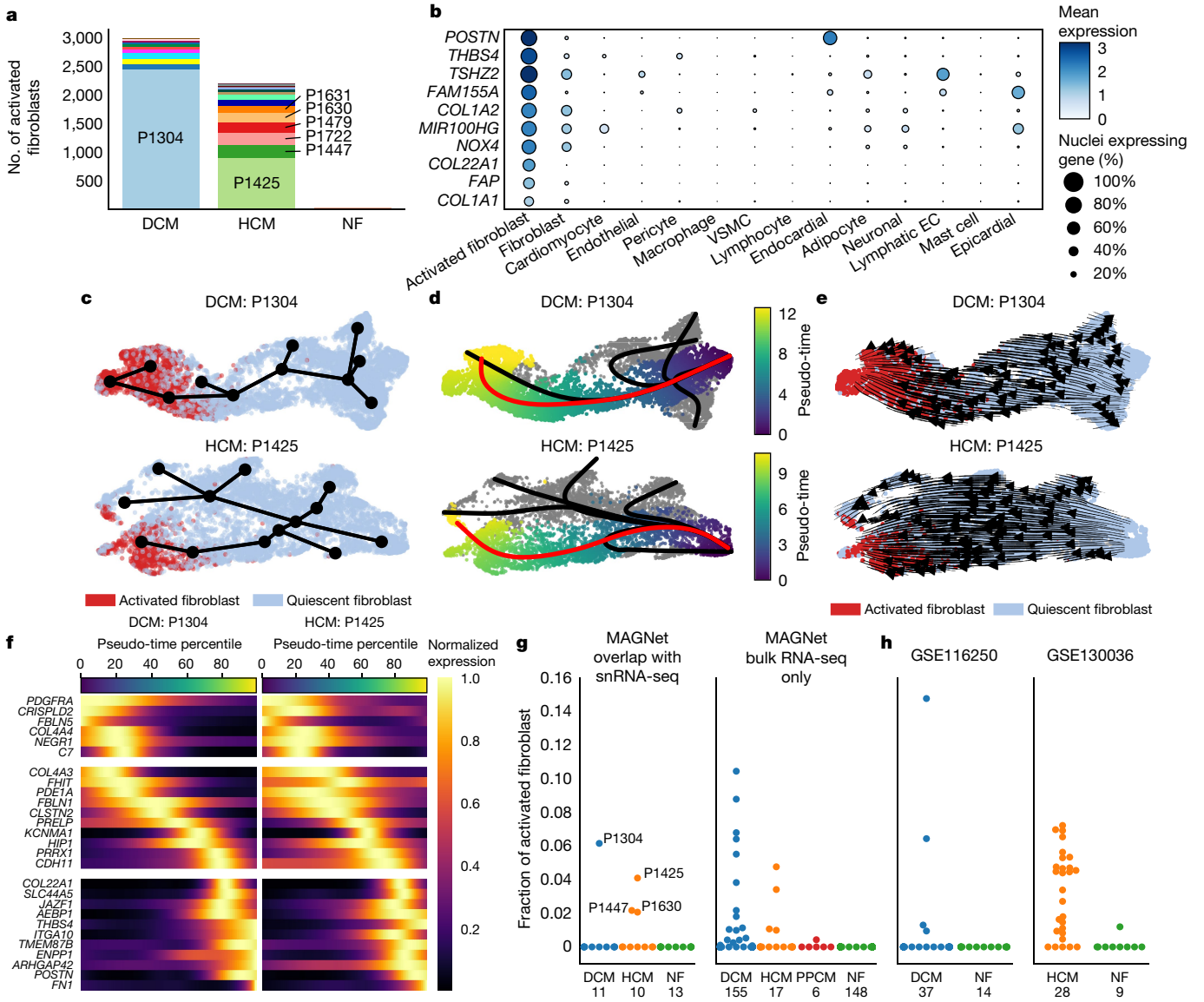


Fig. 3 | DCM- and HCM-specific activated fibroblast populations. **a**, The number of activated fibroblasts per disease state, coloured by donor. **b**, Expression profiles of activated fibroblast marker genes ($n = 42$ donors, 592,689 nuclei). Expression is presented as mean log-normalized expression. **c**, UMAP representation of all fibroblasts from two patients (P1304: $n = 8{,}798$ fibroblasts; P1425: $n = 7{,}164$ fibroblasts) with the largest activated fibroblast populations. Connectivity induced by the minimum spanning tree from Slingshot is overlaid. **d**, UMAP representation with overlaid Slingshot-inferred trajectories coloured by pseudo-time. The highlighted trajectory represents the most connected transition from quiescent to activated fibroblast. **e**, UMAP representation with inferred RNA velocity overlaid as a stream plot. **f**, Predicted expression of genes showing interesting patterns based on a negative binomial generalized additive model (NB-GAM) for each gene across pseudo-time in patient P1304 (left) and patient P1425 (right). Normalized expression is smoothed expression from NB-GAM scaled to the maximum value for each gene. **g**, Estimated fraction of activated fibroblasts from deconvolution analysis of bulk RNA-sequencing data for overlapping data from snRNA-seq (left) and non-overlapping data (right) from the MAGNet study. The number of individuals for each disease state is shown under the respective *x*-axis label. **h**, Deconvolution of bulk RNA-sequencing data from two external datasets displayed as in **g**. PPCM, peripartum cardiomyopathy; EC, Endothelial cell.

macrophages[15] (Extended Data Fig. 8g–i and Supplementary Table 11). This observation is in agreement with the notion that replenishment of CCR2 negative macrophages relies on local proliferation rather than monocyte recruitment.

Activated fibroblasts in DCM and HCM

The starkest change in cell-type composition observed in cardiomyopathy was a cluster of activated fibroblast nuclei that was found almost exclusively in patients with cardiomyopathy (2,989 DCM, 2,196 HCM and 25 NF nuclei), largely derived from DCM patient P1304 and HCM patient P1425 (Fig. 3a). This population showed increased expression of known activated fibroblast markers such as *POSTN*, *NOX4*, *FAP*, *COL1A1* and *COL1A2*[16–19] as well as previously unreported genes such as family with sequence similarity 155 member A (*FAM155A* (also known as *NALF1*)) and teashirt zinc finger homeobox 2 (*TSHZ2*) (Fig. 3b and Supplementary Table 3). An additional activated fibroblast marker, thrombospondin-4 (*THBS4*), has been shown to be up-regulated in matrifibrocytes[20,21] and involved in myocardial remodelling, interstitial fibrosis and cardiomyopathy[20,22]. To validate the presence of this population in diseased patients, we performed RNA in situ hybridization using the highly specific activated fibroblast marker *COL22A1* (log fold change = 8.14, adjusted $P = 3.96 \times 10^{-112}$) and the general fibroblast marker *DCN* (Supplementary Table 3). Hearts from the two patients with the largest activated

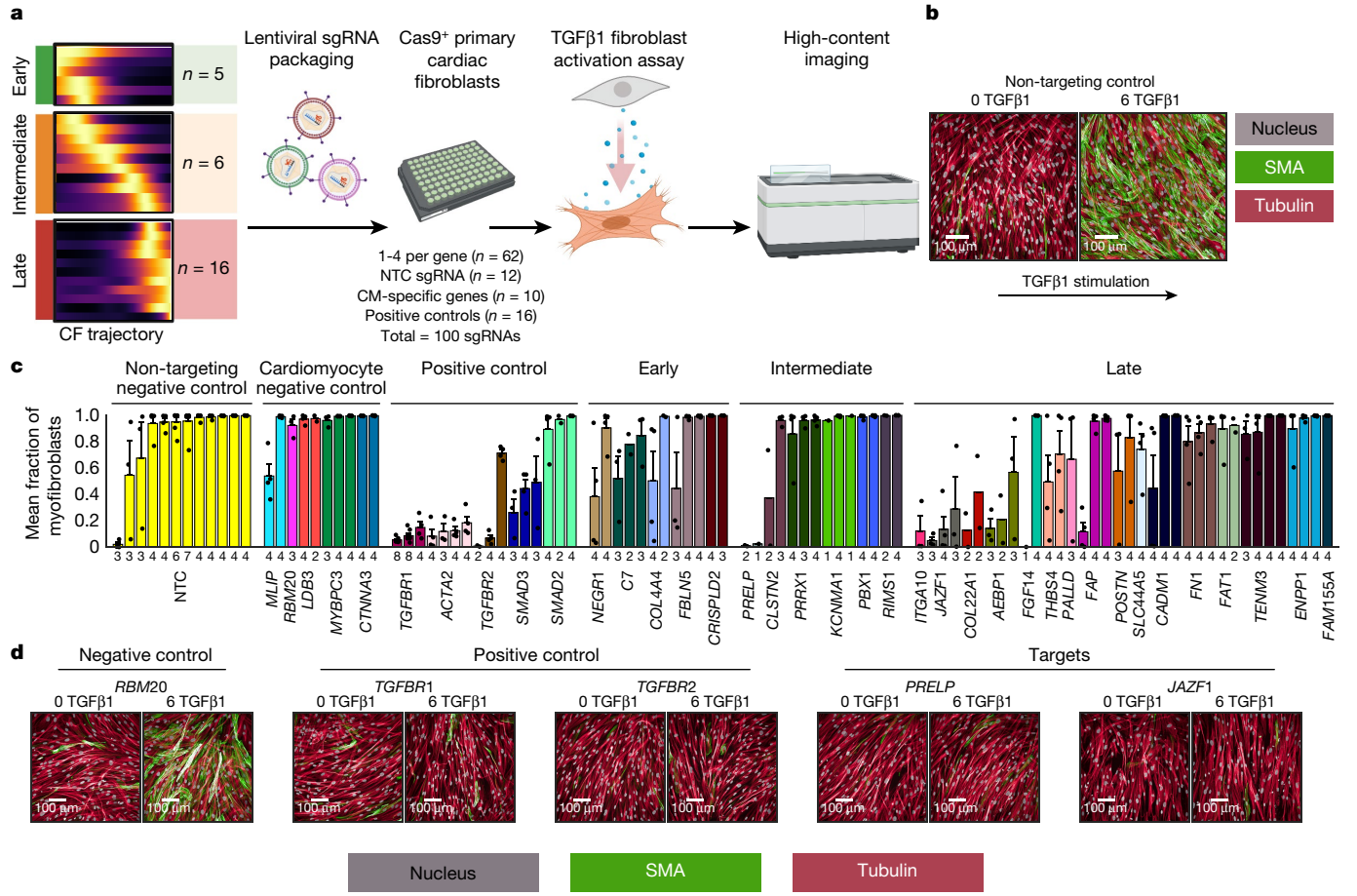


Fig. 4 | Cellular assay of myofibroblast transition in cardiac fibroblasts. **a**, A subset of genes from the activated fibroblast trajectory analysis ($n = 27$) were knocked out in cardiac fibroblasts on 384-well plates using 1–4 sgRNAs per gene, stimulated with TGFβ1, and assessed for changes in cellular phenotypes using high-content imaging. CM, cardiomyocyte. Parts of this figure were created with BioRender.com. **b**, Representative image of cardiac fibroblasts before (left) and after (right) TGFβ1 stimulation with NTC sgRNA. **c**, Bar plots representing the mean fraction of myofibroblasts after TGFβ1 stimulation across all wells for a given sgRNA ($n = 1$–4 wells, with exact number shown under each bar), coloured by gene. Data are mean ± s.e.m. Individual well values for each sgRNA are shown as dots. **d**, Representative images before and after TGFβ1 stimulation for three control genes and two target genes showing strong effects. Details on statistics and reproducibility are provided in Methods.

fibroblast population both had this *COL22A1*-expressing population, which was absent from NF hearts (Extended Data Fig. 9b).

To better understand the gradient of expression between quiescent and activated fibroblasts, we performed a trajectory analysis for the two patients with the largest population of activated fibroblasts, P1304 and P1425 (Fig. 3c–e). Several genes, including *SLC44A5*, *COL22A1*, *POSTN*, *AEBP1*, *JAZF1* and *THBS4* showed increased expression across the trajectory, whereas other genes such as *NEGR1*, *PDGFRA*, *C7*, *FBLN5* and *COL4A4* showed decreased expression across the trajectory (Fig. 3f and Supplementary Table 12). Notably, a subset of genes including *KCNMA1*, *HIP1*, *PRRX1* and *PRELP* showed increased expression at various stages along the gradient (Fig. 3f and Supplementary Table 12). Whether stable intermediate states exist along this trajectory will require further evaluation.

Validation of activated fibroblasts

To examine whether the presence of activated fibroblasts in patients with cardiomyopathy is generalizable to the broader population of patients, we turned to a wider set of patients with DCM, HCM or peripartum cardiomyopathy and NF controls from the Myocardial Applied Genomics Network (MAGNet) study with available bulk RNA-sequencing data. We used cell-type-specific expression profiles derived from our snRNA-seq data to deconvolute bulk RNA-sequencing data and estimate a fraction of activated fibroblasts per patient. In 34 samples with both snRNA-seq and bulk RNA sequencing, deconvolution analysis predicted that 4 samples had a population of activated fibroblasts, 3 of which had the largest populations according to the snRNA-seq data (Fig. 3g and Supplementary Table 13).

Among the 320 independent patients with bulk RNA-sequencing data, 17 out of 155 with DCM, 4 out of 17 with HCM, and none of the 148 NF controls were predicted to have activated fibroblasts, confirming the specificity of this population in patients with cardiomyopathy (Fig. 3g). For a subset of cardiomyopathy patients with available clinical data, we compared those predicted to have an activated fibroblast population to those who were not and found no significant difference in LVEF ($P = 0.42$), LV end diastolic dimension ($P = 0.30$), LV end systolic dimension ($P = 0.38$), LV mass ($P = 0.13$) or posterior wall thickness ($P = 0.41$). Further, we saw no association with a history of diabetes ($P = 0.92$), history of atrial fibrillation ($P = 0.22$) or history of hypertension ($P = 0.96$). However, this may be a result of the currently limited sample size.

We repeated the deconvolution analysis in an independent collection of DCM and NF samples[6] and predicted an activated fibroblast population in 4 out of 37 DCM samples and 0 out of 14 NF samples (Fig. 3h). In a distinct collection of patients with HCM undergoing septal myectomy and NF control samples[23], we predicted an activated fibroblast population in 23 out of 28 HCM samples and 1 out of 9 NF samples (Fig. 3h). The increased prevalence of the activated fibroblast population earlier

in the time course of the disease may suggest a pathological role for these cells, but these findings will require further study.

To validate the deconvolution analysis, we performed RNA in situ hybridization using the activated fibroblast marker gene *COL22A1*. Two patients predicted to have an activated fibroblast population showed increased levels of *COL22A1*, three out of four patients predicted not to have activated fibroblasts showed little to no *COL22A1* expression, and one patient with DCM (P1097) with predicted depletion of activated fibroblasts showed modest expression of *COL22A1* (Extended Data Fig. 10a). In addition, we performed snRNA-seq on three of these patients with cardiomyopathy and confirmed a strong enrichment of activated fibroblasts in one sample in which we predicted an appreciable population of activated fibroblasts using our deconvolution analysis (Extended Data Fig. 10b–d). In sum, our data suggest that this population of activated fibroblasts is almost completely absent from NF hearts, and variably expressed in a subset of patients with DCM and HCM.

Cardiac fibrosis assay

Finally, we developed an assay to measure the progression of cardiac fibroblasts to myofibroblasts upon TGFβ1 stimulation (Fig. 4a). In brief, Cas9-expressing cardiac fibroblasts were plated on 384-well plates and 27 genes identified in our activated fibroblast trajectory analysis were knocked out using 1–4 single guide RNAs (sgRNAs) per gene, with each sgRNA replicated across 1–4 wells after filtering (Supplementary Table 14). As expected, cardiac fibroblasts treated with TGFβ1 displayed a stark phenotypic change in smooth muscle actin (SMA), which we used to define our myofibroblast readout (Fig. 4b). As a positive control, we knocked out several genes known to be involved in TGFβ signalling, including *ACTA2*, *TGFBR1*, *TGFBR2*, *SMAD2* and *SMAD3*. We observed a large reduction in the myofibroblast readout for all *ACTA2* and *TGFBR1* sgRNAs, two out of three *TGFBR2* sgRNAs, and a smaller reduction for *SMAD3* sgRNAs (Fig. 4c). As negative controls, we included 12 non-targeting control (NTC) sgRNAs and 10 sgRNAs targeting 5 cardiomyocyte-specific genes. Nine of the 12 NTC sgRNAs and all but one sgRNA for the 5 cardiomyocyte-specific genes showed no reduction in the myofibroblast readout after TGFβ1 stimulation (Fig. 4c). At least one sgRNA for 1 gene upregulated in intermediate trajectory fibroblasts and 6 genes upregulated in late-stage activated fibroblasts showed a stark reduction (more than 75%) in the myofibroblast readout (Fig. 4c). Each gene from this primary screen was repeated in up to four additional screens to assess consistency of results, and several sgRNAs showed a consistent large effect across multiple screens, including *PRELP*, *JAZF1*, *COL22A1* and *AEBP1* (Extended Data Fig. 11). *PRELP*, *COL22A1* and *AEBP1* exhibited fibroblast-specific expression, whereas *JAZF1* also showed appreciable expression in cardiomyocytes (Extended Data Fig. 12). However, any disease-specific changes in expression were restricted to fibroblast cell types (Supplementary Table 15). As examples of sgRNAs inhibiting the transition to the myofibroblast state, we observed a substantial reduction in SMA for *PRELP* and *JAZF1* compared with positive and negative controls in our primary screen (Fig. 4d). We then individually cloned the top-performing guides into a LentiCRISPR v2 backbone, transduced primary cardiac fibroblasts, and confirmed the knockout efficiency by real-time PCR (Supplementary Table 16 and Methods).

Discussion

In our analyses, we identified large transcriptional changes at single-cell resolution between LV from patients with DCM or HCM compared with NF donors, with relatively few differences between DCM and HCM LV. Although there are clear differences between the genetic bases of DCM and HCM, our findings suggest a convergence to a common transcriptional profile in advanced cardiomyopathy requiring transplantation. Moreover, we uncovered compositional shifts in cell types or cell sub-types between cardiomyopathy and NF hearts. The most prominent shifts included a reduction of proliferating CCR2-negative macrophages in cardiomyopathy hearts and an increase in activated fibroblasts in cardiomyopathy hearts. Activated fibroblasts were found almost exclusively in patients with DCM or HCM and were computationally predicted to be present in an independent subset of patients with cardiomyopathy.

To expand our understanding of cardiac fibroblast activation, our snRNA-seq data identified several genes whose expression changed along the quiescent-to-activated fibroblast trajectory that also showed a reduction in the myofibroblast readout when knocked out in our cellular assay of TGFβ1-induced fibroblast activation. One of the most prominent of these genes, *PRELP*, which encodes the extracellular matrix protein prolargin, may serve as a basement membrane anchor and is known to bind collagen type I and type II, suggesting a plausible role in cardiac fibrosis[24]. A second gene, *JAZF1*—which encodes juxtaposed with another zinc finger protein 1—has largely been studied in the context of type 2 diabetes and has a role in lipid and glucose metabolism as well as insulin sensitivity[25–27]. A role in cardiac fibrosis is currently unexplored. *COL22A1*—which encodes the collagen type XXII α1 chain—is expressed at myotendinous junctions, particularly in skeletal and heart muscle[28]. In dermal fibroblasts, silencing *COL22A1* impeded the myofibroblast transition, as indicated by decreased expression of *ACTA2*, consistent with our observation in cardiac fibroblasts[29].

Our study is not without limitations. First, we have focused on advanced cardiomyopathy requiring transplantation. In future work, it would be useful to determine whether there are more overt differences in DCM and HCM earlier in the course of disease, but such studies will probably be limited by sample availability. Second, our study included only individuals with European ancestry, and expanding to other populations will be of interest. Third, estimating a fraction of activated fibroblasts per patient from snRNA-seq and deconvolution is not without measurement error. In the former, there is probably some degree of nucleus misclassification on the border of the quiescent and activated fibroblast populations. In the latter, computational algorithms for deconvolution have inherent error rates and may not decouple rare cell types, particularly when they share expression features of other cell types. Third, although the quiescent-to-activated fibroblast trajectories were constructed using only two individuals, we note that several thousand individual nuclei from our discovery and validation samples had a similar expression pattern. Finally, although we observed clear evidence that knockout of a subset of genes identified from our snRNA-seq analysis inhibits fibroblast activation, we currently do not know how well the myofibroblast transition from TGFβ1 in cultured cardiac fibroblasts reflects our snRNA-seq inferred trajectory.

In conclusion, we have comprehensively characterized the transcriptional landscape at the cellular level from the human heart in health and in patients with DCM or HCM. Our results identify both a final common transcriptional pathway and reduced numbers of proliferating resident cardiac macrophages in patients with cardiomyopathy. Finally, we provide a comprehensive transcriptional profile of a disease-specific activated fibroblast population and suggest mechanistic roles in fibroblast activation for several genes, expanding the current understanding of these cells in fibrotic disease of the heart. In aggregate, our findings extend our understanding of the transcriptional and molecular basis of cardiomyopathies, results that will further inform the pathways and potential therapeutic targets for these morbid cardiac conditions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04817-8.

1. Savarese, G. & Lund, L. H. Global Public Health Burden of Heart Failure. *Card. Fail. Rev.* **03**, 7 (2017).
2. Oktay, A. A. et al. *Diabetes, Cardiomyopathy, and Heart Failure. Endotext* (MDText.com, Inc., 2000).
3. Liu, Y. et al. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* **105**, 83–89 (2015).
4. Chen, C. Y. et al. Suppression of detyrosinated microtubules improves cardiomyocyte function in human heart failure. *Nat. Med.* **24**, 1225–1233 (2018).
5. Maron, B. J. et al. Contemporary definitions and classification of the cardiomyopathies: An American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. *Circulation* **113**, 1807–1816 (2006).
6. Sweet, M. E. et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. *BMC Genomics* **19**, 812 (2018).
7. Tucker, N. R. et al. Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).
8. Litviňuková, M. et al. Cells of the adult human heart. *Nature* **588**, 466–472 (2020).
9. Wang, L. et al. Single-cell reconstruction of the adult human heart during heart failure and recovery reveals the cellular landscape underlying cardiac function. *Nat. Cell Biol.* **22**, 108–119 (2020).
10. Ashburner, M. et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
11. Carbon, S. et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
12. Kalucka, J. et al. Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764–779.e20 (2020).
13. Crinier, A. et al. High-Dimensional Single-Cell Analysis Identifies Organ-Specific Signatures and Conserved NK Cell Subsets in Humans and Mice. *Immunity* **49**, 971–986.e5 (2018).
14. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
15. Bajpai, G. et al. The human heart contains distinct macrophage subsets with divergent origins and functions. *Nat. Med.* **24**, 1234–1245 (2018).
16. Tallquist, M. D. & Molkentin, J. D. Redefining the identity of cardiac fibroblasts. *Nat. Rev. Cardiol.* **14**, 484–491 (2017).
17. Cucoranu, I. et al. NAD(P)H oxidase 4 mediates transforming growth factor-β1-induced differentiation of cardiac fibroblasts into myofibroblasts. *Circ. Res.* **97**, 900–907 (2005).
18. Tillmanns, J. et al. Fibroblast activation protein alpha expression identifies activated fibroblasts after myocardial infarction. *J. Mol. Cell. Cardiol.* **87**, 194–203 (2015).
19. Shinde, A. V. & Frangogiannis, N. G. Mechanisms of Fibroblast Activation in the Remodeling Myocardium. *Curr. Pathobiol. Rep.* **5**, 145–152 (2017).
20. McLellan, M. A. et al. High-Resolution Transcriptomic Profiling of the Heart During Chronic Stress Reveals Cellular Drivers of Cardiac Fibrosis and Hypertrophy. *Circulation* **142**, 1448–1463 (2020).
21. Forte, E. et al. Dynamic Interstitial Cell Response during Myocardial Infarction Predicts Resilience to Rupture in Genetically Diverse Mice. *Cell Rep.* **30**, 3149–3163 (2020).
22. Frolova, E. G. et al. Thrombospondin-4 regulates fibrosis and remodeling of the myocardium in response to pressure overload. *FASEB J.* **26**, 2363–2373 (2012).
23. Liu, X. et al. Long non-coding and coding RNA profiling using strand-specific RNA-seq in human hypertrophic cardiomyopathy. *Sci. Data* **6**, 90 (2019).
24. Bengtsson, E. et al. The Leucine-rich Repeat Protein PRELP Binds Perlecan and Collagens and May Function as a Basement Membrane Anchor. *J. Biol. Chem.* **277**, 15061–15068 (2002).
25. Li, L. et al. The role of JAZF1 on lipid metabolism and related genes in vitro. *Metabolism.* **60**, 523–530 (2011).
26. Guang-feng, M. et al. JAZF1 can regulate the expression of lipid metabolic genes and inhibit lipid accumulation in adipocytes. *Biochem. Biophys. Res. Commun.* **445**, 673–680 (2014).
27. Yuan, L. et al. Transcription factor TIP27 regulates glucose homeostasis and insulin sensitivity in a PI3-kinase/Akt-dependent manner in mice. *Int. J. Obes.* **39**, 949–958 (2015).
28. Koch, M. et al. A Novel Marker of Tissue Junctions, Collagen XXII. *J. Biol. Chem.* **279**, 22514 (2004).
29. Watanabe, T. et al. A Human Skin Model Recapitulates Systemic Sclerosis Dermal Fibrosis and Identifies COL22A1 as a TGFβ Early Response Gene that Mediates Fibroblast to Myofibroblast Transition. *Genes (Basel).* **10**, 75 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Human tissue samples

Adult human myocardial samples of European ancestry were collected from organ donors by the Myocardial Applied Genetics Network (MAGNet; www.med.upenn.edu/magnet) as previously described[7,30]. In brief, NF samples were obtained from organ donors with no history of heart failure and samples from failing human hearts were obtained from explanted hearts of donors receiving a heart transplant. Hearts were arrested in situ with at least 1 l ice-cold crystalloid cardioplegia solution and were transported to the laboratory in ice cold cardioplegia solution until cryopreservation (always <4 h). Transmural cardiac tissue was taken from the left ventricular free wall of a non-infarcted region excluding the septum, and predominantly comprised the anterior wall midway between the apex and base. Written informed consent for research use of donated tissue was obtained from next of kin in all cases. Research use of tissues were approved by the relevant institutional review boards at the Gift-of-Life Donor Program, the University of Pennsylvania, Massachusetts General Hospital and the Broad Institute.

snRNA-seq

Single-nucleus suspensions were generated as previously described[7]. In brief, left ventricular tissue (approximately 100 mg) was sectioned at 100 μm (Leica CM1950 cryostat), resuspended in 4 ml ice cold lysis buffer (250 mM sucrose, 25 mM KCl, 0.05% IGEPAL-630, 3 mM $MgCl_2$, 1 μM DTT, 10 mM Tris pH 8.0) and homogenized using a dounce. Large debris was pelleted at 40*g* for 1 min (Beckman Coulter Allegra X-15R swinging bucket centrifuge) and the supernatant was filtered sequentially through a 50 μm and 10 μm filter (pluriSelect Life Science). The filters were washed with 6 ml PBS wash buffer (0.01% BSA, 5 mM $MgCl_2$, PBS) and nuclei were pelleted at 550*g* for 5 min at 4 °C, washed in 6 ml of nuclear wash buffer and recentrifuged. After removal of wash buffer, nuclei were resuspended in approximately 150 μl of cold nuclear resuspension buffer (Nuclear wash buffer + 0.4 U μl^{-1} murine RNAse inhibitor (New England Biolabs)) with gentle trituration, mixed 1:1 with trypan blue, then counted on a haemocytometer. Cells were loaded into the 10x Genomics microfluidic platform (Single cell 3′ solution, v3) for an estimated recovery of 5,000 cells per device. Processing of libraries was performed according to manufacturer's instructions with a few modifications. First, nuclei were incubated at 4 °C for 15 min after emulsion generation to promote nuclear lysis. Second, the reverse transcription protocol was modified to be 42 °C for 20 min then 53 °C for 120 min. Libraries were multiplexed at an average of four libraries per flow cell on an Illumina Nextseq550 in the Broad Institute's Genomics Platform.

snRNA-seq processing

Raw base call files for each sample were de-multiplexed and converted to FASTQ files using the 10x Genomics toolkit CellRanger 4.0.0 (cellranger mkfastq). Prior to count quantification with CellRanger, we trimmed homopolymer repeats (A_{30}, C_{30}, G_{30} and T_{30}) as well as the template-switch oligonucleotide and its reverse complement (AAGCAGTGGTATCAACGCAGAGTACATGGG and CCCATGTACTCTG CGTTGATACCACTGCTT) using cutadapt v1.18[31] with parameters (max_error_rate = 0.1, min_overlap = 20) and (max_error_rate = 0.07, min_overlap = 10), respectively. Trimmed reads were aligned to the GRCh38 pre-mRNA human reference (v2020-A) using CellRanger 4.0.0 (cellranger count) with --expect-cells 5000. All other parameters were left as defaults.

Quality control metrics from cellranger count were inspected for each sample (Extended Data Fig. 1a and Supplementary Table 2). Six samples with <50% of reads in cells, <65% of reads confidently mapping to transcriptome, <90% valid barcodes, or abnormally low Q30 were excluded from downstream analysis. Two additional samples were removed after visually inspecting unique molecular identifier (UMI) decay curves and noting a lack of sufficient ambient plateau for subsequent cell calling and background removal with CellBender remove-background v0.2[32] (https://github.com/broadinstitute/CellBender) (Extended Data Fig. 1b). Finally, the distribution of counts from Y-chromosome genes were compared between phenotypically classified males and females. No outliers were identified (Extended Data Fig. 1c).

Samples passing quality control were processed with the remove-background tool from CellBender v0.2[32] to both call non-empty droplets and subtract ambient background RNA contamination from count matrices. The following parameters were used: --expected-cells 5000, --z-dim 100, --total-droplets-included 25000, --epochs 150, --learning_rate 1e-4, and --fpr 0.01. For 3 of the 80 samples (LV_1539_1, LV_1549_1, and LV_1735_1), --learning_rate was adjusted to 5e-5 to reach training convergence.

We applied the tool scR-Invex (https://github.com/broadinstitute/scrinvex) to assign reads for each sample to exons, introns, or spanning both. For each cell barcode, we calculated the proportion of total exonic reads in the droplet. As we are performing snRNA-seq, an increased proportion of exonic reads may indicate a larger component of cytoplasmic material for a given droplet.

Nuclei map construction

Nuclei deemed non-empty based on CellBender v0.2 (885,944 nuclei) were jointly aggregated. All data processing was doing in scanpy v1.6.0[33] unless otherwise noted. First, the top 2,000 most highly variable genes were selected using sc.pp.highly_variable_genes (flavor = 'seurat_v3'). Count data was then log-normalized by nucleus using sc.pp.normalize(1e4) followed by sc.pp.log1p(). Expression of highly variable genes was scaled to unit variance and zero mean with sc.pp.scale(). The top 50 principal components were then estimated using sc.tl.pca(). To account for large differences in expression across patients as well as technical batch effects, principal components were aligned with Harmony[34] as implemented in harmony-pytorch v0.1.4 (https://github.com/lilab-bcb/harmony-pytorch) using patient as a batch indicator. A neighborhood graph was constructed using all 50 adjusted principal components with Euclidean distance sc.pp.neighbors(n_neighbors = 15). To visualize clusters, UMAP was run on the resulting neighborhood, sc.tl.umap(min_dist = 0.2). Leiden clustering at resolution 1.5 was used to cluster nuclei into 47 groups, sc.tl.leiden(resolution = 1.5).

Key quality control metrics were calculated for all nuclei including the percent of reads mapped to mitochondrial genes based on CellRanger counts, the proportion of reads mapping to purely exonic regions, entropy as calculated using Bayesian entropy estimation from the ndd python library (https://pypi.org/project/ndd/1.6.3/), and a predicted doublet score as estimated using Scublet[35] per sample based on CellRanger counts. Seven clusters comprised of a total 63,068 droplets were identified as enriched for mitochondrial reads and/or the proportion of exonic reads and depleted for entropy (Extended Data Fig. 2). Eleven clusters comprised of 77,534 droplets were identified as enriched for doublet populations (Extended Data Fig. 2). All nuclei from these clusters were subsequently removed.

Additionally, we identified low-quality nuclei within each sample and cluster combination. Performing this quality control by sample and cluster is important to account for heterogeneity between sample preparations and genuine biological differences in complexity by cell type. In brief, for each combination of sample and cluster, an upper and lower bound were set at the 75th percentile plus 1.5 times the interquartile range (IQR) and the 25th percentile minus 1.5 times IQR, respectively, for total UMI, number of genes (n_gene), entropy, and log(n_gene) × entropy. An upper bound for percent of reads mapped to mitochondrial genes, proportion of exonic reads, and doublet score from Scrublet was set at

the 75th percentile plus 1.5 times IQR. If a given sample and cluster combination had fewer than 30 nuclei, distributional outliers were unreliable and therefore hard thresholds for removal were set as total UMI > 15,000, n_gene > 6,000, entropy < 8, proportion of exonic reads > 0.18, doublet score > 0.30, and log(n_gene) × entropy > 75. A final hard cut-off was set to exclude any nuclei with less than 150 UMI, less than 150 genes, or greater than 5% of reads mapping to mitochondrial genes. In total, this removed 280,630 droplets (Extended Data Fig. 2). Finally, the remaining 605,314 nuclei were re-clustered as previously described using cosine distance, increasing the number of principal components to 100, and applying Leiden clustering at a resolution of 0.6.

Sub-clustering was performed on each cell type to identify remaining low-quality or misclassified nuclei in each cell type. Sub-clustering was performed using an iterative approach on major cell types. For each subset, a new neighborhood graph was constructed based on the global Harmony-adjusted principal components. Then, Leiden resolution was increased from 0.05 to 1.0 in increments of 0.05 until a cluster emerged with no genes having area under the receiver operating characteristic curve (AUC) > 0.60 in predicting that class. Sub-clusters of misclassified nuclei were identified by scoring nuclei with sc.tl.score_genes()[14] based on the top 50 markers, as ranked by AUC, of every cell type as well as all mitochondrial genes. Sub-clusters enriched for scores of other cell types were considered misclassified or low quality. Based on elevated scores, we identified 12,625 out of 605,314 (2.1%) nuclei as potentially low quality or misclassified and removed them from subsequent analysis (Extended Data Fig. 2).

Cell-type identification

A list of genes selectively expressed in each cell type was constructed using a combination of strategies. First, we estimated an AUC at a nuclei level, agnostic of sample by classifying each nucleus as either coming from the target cell type or not, and then predicting this class with the normalized expression for each gene. Second, a formal differential expression model was run on the aggregation of gene counts for each patient as motivated by Lun and Marioni, 2017[36]. For each cell type, gene counts for all nuclei belonging to either technical replicate from a patient were summed together if the total nuclei counts was greater than 50. Genes were filtered out if their counts were too low for testing as determined by the edgeR function filterByExpr(group=cell type)[37]. DESeq2 normalization[38] was applied to estimate scaling factors by sample/cell-type combination and then analyzed using the limma–voom pipeline[39] with a design of '0 + cell type + individual'. Contrasts were fit to extract differential expression results for each cell type and two-sided *P*-values were adjusted with a Benjamini–Hochberg correction.

Marker genes were selected as protein coding genes expressed in at least 25% of nuclei from the target cell type of interest, with an AUC for the target cell type greater than 0.60, a log-fold change from the limma–voom model greater than 2, and an FDR adjusted *P*-value from the limma–voom model <0.01. Enrichment for biological process gene ontologies for each set of marker genes was performed using the R package GOstats version 2.46.0 with one-sided *P*-values[40]. Only gene ontologies with more than 10 genes, but less than 1,000 genes, were considered. A significance threshold was set at $0.05/54{,}009 = 9.25 \times 10^{-7}$ to account for 54,009 ontologies tested across all cell types. Marker genes and enriched ontologies were used to assign cell-type labels to each cluster. Cell-type centroids were clustered based on the mean expression of the top 2,000 most highly variable genes using Euclidean distance and Ward linkage to display as a dendrogram.

Composition analysis

All composition analyses were performed using the Bayesian method scCODA 0.1.2.post1[41]. A default Normal prior was used for intercepts, a spike-and-slab prior for each covariate on a cell type, and a non-centered parameterization of a normal distribution prior for significant effects. Hamiltonian Monte Carlo was done with 20,000 iterations, using 5,000 iterations as burn-in. Since compositional analysis is performed relative to a reference cell type, we selected a reference cell type using a method inspired by microbiome work[42]. In brief, for each cell type *i*, we calculated the standard deviation of the $\log_{10}$ ratio of cell type *i* counts to cell type *j* across samples, for each other cell type *j*. We took the median of those standard deviations and found the cell type with the lowest median and assigned that as the reference group. All composition analyses were adjusted for sex.

Differential expression testing

Differential expression testing was performed by cell type. To account for the correlation in expression among nuclei from a given individual, we summed counts for genes across nuclei for each patient within each cell type, requiring a minimum of 25 nuclei. We applied a limma–voom model adjusting for age and sex using the model Gene_Expression ~ group + sex + age for each gene. We excluded mitochondrial genes and ribosomal genes, removed genes in less than 1% of nuclei in the given cell type of both groups being compared, and applied an additional filter for lowly expressed genes using the filterByExpr(group=group)[37] function. Two-sided *P*-values were calculated. A Benjamini–Hochberg correction was applied for multiple testing correction and an FDR-adjusted *P*-value < 0.01 was used for significance. Differential expression tests were run on both CellBender remove-background and CellRanger counts. We focus our interpretation of genes found significantly differentially expressed using both the CellBender and CellRanger counts to ensure robustness in the data presented.

To identify genes with a different effect between cardiomyopathy and NF donors dependent on sex, a second differential expression test was performed including an interaction term for sex × disease. Due to limited statistical power given the low sample size, a more lenient FDR was set to 0.1 for exploratory analysis.

Ambient, background contamination flag

Generally, in snRNA-seq experiments, some ambient background RNA remains in most droplets despite running CellBender remove-background. To flag genes as having a high probability of coming from background we calculated a heuristic by the following. First, for each gene we calculated two positive predictive values for each cell type by dichotomizing expression as 0 or > 0 (PPV0) and ≤ 1 or > 1 (PPV1) and predicting cell-type class. Second, we calculated the mean of PPV0 and PPV1. We then flagged genes as potentially coming from background if 1. The maximum mean(PPV0, PPV1) across all cell types minus the mean(PPV0, PPV1) for the given cell type is greater than 0.5 and 2. The average log-normalized expression in the cell type with the maximum mean(PPV0, PPV1) is greater than the cell type of interest.

Sample PCA

PCA was performed on counts summed across nuclei in a patient. Lowly expressed genes, identified as those with less than 10 counts in total, were removed prior to analysis. A variance stabilizing transformation (vst) was applied to counts after estimating the dispersion–mean relation using the R package DESeq2[38]. Principal components were estimated using the *prcomp* function in R v3.5.0 using the top 500 highly variable genes.

Reactome pathway enrichment

After filtering out genes with expression potentially driven by background contamination, we performed an enrichment test of Reactome pathways[43] by cell type and disease comparison using two approaches. First, we performed a GSEA using fgsea v1.16.0[44] on gene sets with greater than 15 genes but less than 500 genes. We sorted genes by the *t*-statistic from the differential expression test using CellBender count data and tested for enrichment in Reactome pathways from MsigDB (https://www.gsea-msigdb.org/gsea/msigdb). Second, we performed a hypergeometric test of enrichment using ReactomePA to emphasize extreme

changes[45]. We selected genes with a Benjamini–Hochberg corrected *P*-value <0.01 based on both CellRanger and CellBender counts and split these into upregulated genes (greater than top 90th percentile of the absolute value of all log fold-changes) for the given cell type or downregulated genes (less than −1 × top 90th percentile of the absolute value log fold-changes). Given the number of differentially expressed genes, this second approach allows us to emphasize more extreme differential expression results while also providing a control for large discrepancies between CellBender and CellRanger tests. Pathways enriched in both tests (Benjamini–Hochberg corrected two-sided *P*-value < 0.05 for fgsea and Benjamini–Hochberg corrected one-sided *P*-value < 0.05 for ReactomePA) with consistent directionality were considered robust results.

Sub-clustering analysis

A targeted sub-clustering analysis was performed within the most abundant cell types: cardiomyocytes, fibroblasts, endothelial cells, pericytes, macrophages, VSMCs, and lymphocytes. Clustering was performed as was done in the global map with a reduction to n_neighbors = 10 in sc.pp.neighbors to capture more local structure. Additionally, we regressed out both the percent of mitochondrial reads and total UMI per nucleus prior to PCA calculation. Leiden clustering within each cell type was performed at resolutions between 0.1 and 1.0, at increments at 0.1. When a sub-cluster emerged that had no marker genes with AUC > 0.6, clustering was halted, and the previous resolution was retained. In the case of VSMCs, we altered these criteria to require at least 2 genes with AUC > 0.6 to avoid over-clustering. Marker genes were calculated analogously to the global cluster analysis. We used more liberal criteria to define marker genes requiring the gene to be expressed in at least 15% of nuclei in the target sub-cluster, an AUC greater than 0.55, a log fold-change greater than 0, and an FDR-adjusted two-sided *P*-value < 0.01. Composition analysis was performed using scCODA[41] as described previously to identify subpopulations with varying abundance in disease states. For macrophages, nuclei were scored for cell-cycle profiles using sc.tl.score_genes_cell_cycle() with default settings.

Activated fibroblast trajectory analysis

An appreciable activated fibroblast population (>500 nuclei) was identified in 2 patients (P1304 and P1425). We extracted 8,798 and 7,164 fibroblast nuclei for patient P1304 and patient P1425, respectively, and inferred cell-state trajectories between quiescent and activated states within each sample using slingshot 1.6.0[46] in R v3.6.0. Genes with non-zero counts in less than 20 nuclei were removed and the remaining data were normalized with Seurat v3.1.5 using SCTransform[47], regressing out technical effects from the total number of UMI and the percent of mitochondrial reads. We performed PCA on the top 3,000 highly variable genes and ran UMAP on the top 30 principal components using cosine distance, min.dist = 0.1, and n.neighbors = 15. A low min.dist was set to preserve local structure more accurately. We next calculated a neighborhood graph (FindNeighbors) followed by Leiden clustering (FindClusters) at a resolution of 1.0. Slingshot was run on this UMAP space based on the resulting clustering. The most representative trajectory from quiescent to activated fibroblasts was selected by following the most connected pathway across the minimum spanning tree inferred by slingshot.

In addition to slingshot, we inferred RNA velocity in these two patients using velocyto 0.17.17[48] with default settings followed by scVelo 0.2.3[49]. As we have performed snRNA-seq, we were limited to 17% spliced reads in P1304 and 16% spliced reads in P1425. We filtered out genes with less than 20 spliced and unspliced counts and retained the top 3,500 most variable genes. We computed the first and second-order moments for each cell based on 30 nearest neighbors, obtaining the neighborhood graph from Euclidean distance of the top 30 principal components from PCA. We then ran the dynamical model with default settings.

To identify interesting gene expression patterns along the slingshot inferred trajectories, we applied the method tradeSeq in R v3.6.0[50]. A generalized additive model (GAM) was fit for each gene using fitGAM(nknots = 6) on SCTransform adjusted counts. We looked for genes that showed associations with slingshot-inferred pseudo-time using both startVsEndTest() and associationTest() in tradeSeq. Strong markers of the early trajectory were identified as those with an FDR adjusted two-sided *P*-value < 0.05 from startVsEndTest(), log fold-change < −1, and found in at least 30% of nuclei at the earliest cluster of the trajectory. Strong markers of the late trajectory were identified as those with an FDR adjusted two-sided *P*-value < 0.05 from startVsEndTest(), log fold-change > 1 and found in at least 30% of the latest cluster of the trajectory. Genes showing other interesting patterns were identified as having an FDR adjusted two-sided *P*-value < 0.05 in associationTest() and being most highly expressed in a cluster other than the start or end cluster with a minimum of 30% of nuclei with non-zero expression. We then subset to genes that showed consistent patterns across P1304 and P1425 for visualization.

Activated fibroblast deconvolution

We applied CIBERSORTx[51] (https://cibersortx.stanford.edu/) to computationally deconvolute bulk RNA-sequencing experiments and estimate a fraction of activated fibroblasts. Cell-type signatures were derived from our global map after collapsing subgroups into broad cell types. Markers for each cell type were selected as those expressed in at least 30% of the target cluster at non-zero levels, AUC > 0.60, and significant in the limma–voom model with FDR < 0.01 and log fold-change > 2. A maximum of 200 genes was selected for each cell type, retaining those with the highest log fold-change estimate.

Bulk RNA-sequencing gene abundance data from the MAGNet study containing DCM, HCM, PPCM (peripartum cardiomyopathy) and NF patients was downloaded from the Gene Expression Omnibus (GEO) (accession ID: GSE141910). Publicly available bulk RNA-sequencing data from additional DCM and NF samples was obtained from GEO (accession ID: GSE116250) and the Sequence Read Archive (SRA) (accession ID: PRJNA477855). FASTQs were aligned to GENCODE v36 with salmon v1.4.0[52] to quantify transcripts per million (TPM) as input to CIBERSORTx. Finally, publicly available bulk RNA-sequencing data from HCM and NF patients as reported in Liu et al. (2019)[23] was obtained from GEO using accession number GSE130036 and used as input to CIBERSORTx.

Within the MAGNet study, at least some clinical data was available for 73 patients with DCM and 17 patients with HCM. We tested for differences between cardiomyopathy patients predicted to have an activated fibroblast population (actFB) and those predicted not to have an activated fibroblast population (no_actFB) on the following features: LVEF ($n_{actFB} = 12$; $n_{no_actFB} = 77$), LV end diastolic volume ($n_{actFB} = 11$; $n_{no_actFB} = 60$), LV end systolic volume ($n_{actFB} = 11$; $n_{no_actFB} = 56$), LV mass ($n_{actFB} = 8$; $n_{no_actFB} = 50$), posterior wall thickness ($n_{actFB} = 10$; $n_{no_actFB} = 53$), history of diabetes ($n_{actFB} = 13$; $n_{no_actFB} = 76$), history of atrial fibrillation ($n_{actFB} = 11$; $n_{no_actFB} = 75$), or history of hypertension ($n_{actFB} = 12$; $n_{no_actFB} = 74$). For continuous measures, a two-sided *t*-test was used. For categorical features, a chi-square test was used.

Whole-genome sequencing analysis

Whole-genome sequencing was performed targeting 30× coverage at the Broad Institute of Harvard and the Massachusetts Institute of Technology using the Illumina NovaSeq platform. Reads were aligned to the GRCh38 reference using BWA-MEM[53]. Variants were called using GATK HaplotypeCaller v3.5.0[54]. In total, 10 out of 11 patients with DCM, 15 out of 15 patients with HCM, and 15 out of 16 NF donors were successfully sequenced. All samples had mean coverage >25× and <5% chimeras. Variants in 106 genes used in clinical testing panels for cardiomyopathy (Invitae Cardiomyopathy Comprehensive Panel) were annotated for predicted LOF variation using the LOFTEE plugin (https://github.com/konradjk/loftee) for the Ensembl Variant Effect Predictor[55] in Hail (https://github.com/hail-is/hail). We removed insertions or deletions with quality by depth (QD) ≤ 3 or single-nucleotide polymorphisms with QD ≤ 2, variants with a failing VQSR flag, variants with a LowQual flag,

variants with an InbreedingCoeff flag, variants within low complexity region, and variants in segmental duplications. Only high-confidence LOF variants in canonical transcripts without the LOFTEE flags PHYLOCSF_WEAK or NON_CAN_SPLICE were considered when deeming a sample a LOF carrier. Additional pathogenic variants were identified by looking for pathogenic, or likely pathogenic, missense variants in ClinVar (downloaded on 12 December 2020).

Activated fibroblast RNAscope validation

RNA in situ hybridization was performed as previously described[7] using RNAscope 2.5 duplex assay (ACD bio). Briefly, fresh frozen samples were sectioned at 10 µm and mounted onto Superfrost plus slides (VWR). Samples were fixed in 4% PFA for 30 min at room temp and treated with protease IV for 30 min. The RNAscope assay was then carried out using the manufacturer's protocol (ACD bio). Probes for *Col22A1* (Channel 2) and *DCN* (Channel 1) were purchased from ACD bio and samples were imaged on a Zeiss Observer Z1 microscope.

Macrophage imaging

Left Ventricular sections (10 µm) were fixed (4% PFA, 15 min), permeabilized (0.1% TX-100, 5 min), blocked (7% donkey serum, 1 h) and incubated at 4 °C overnight with primary antibodies Ki67 (ab15580, Abcam) at 1:100 dilution and CD163 (ab156769, Abcam) at 1:100 dilution in 7% donkey serum. The following day, sections were washed 3 times in PBS and incubated with secondary antibodies and DAPI in the dark for 1 h at room temperature including Alexa-fluor donkey anti-mouse 568 (A10037, Thermofisher) at 1:200 dilution and Alexa-fluor donkey anti-rabbit 488 (A32790, Thermofisher) at 1:200 dilution in 7% donkey serum. Samples were mounted with prolong gold (Thermofisher) and representative images were taken at 20× using a Leica SP8 confocal microscope. For cell counting, tiled images were taken of the whole section at 10× and macrophages and Ki67-positive macrophages were counted using Fiji v2.1.0[56].

Cardiac fibrosis assay

Genes included in the cardiac fibrosis assay were selected from the snRNA-seq trajectory analysis as follows. First, 6 early-trajectory, 59 intermediate-trajectory and 26 late-trajectory genes were identified as described in 'Activated fibroblast trajectory analysis'. We further reduced this by performing a Wilcoxon test between subclusters of fibroblasts in DCM P1304 and HCM P1425, separately. Early trajectory genes were selected as those with log fold-change > 1 and FDR < 0.01 comparing the first sub-cluster along the trajectory to all others ($n = 5$). Intermediate trajectory genes were selected as those with log fold-change > 1 and FDR < 0.01 comparing any sub-cluster between the first and last with all others ($n = 9$). Late trajectory genes were selected as those with log fold-change > 1 and FDR < 0.01 comparing the last sub-cluster along the trajectory to all others ($n = 26$). We then filtered 3 intermediate trajectory genes after visual inspection and selected late trajectory genes with strong specificity toward the global activated fibroblast population leaving 5 early, 6 intermediate, and 16 late trajectory genes.

Primary human cardiac fibroblasts (ACBRI5118, Cell Systems) were obtained directly from Cell Systems and no mycoplasma tests were conducted. Stable Cas9-expressing fibroblasts were generated from these via transduction with lentivirus particles containing a constitutively expressed Cas9 followed by blasticidin selection (VCAS10129, Horizon Discovery). The LentiGuide-Puro (Addgene plasmid #52963)[57] and pXPR_050 (Addgene plasmid #96925)[58] lentiviral vectors were used to deliver sgRNAs that were designed using CRISPick[58,59]. Lentiviral sgRNAs were cloned and generated by the Broad Genetic Perturbation Platform (GPP). All sgRNA sequences can be found in Supplementary Table 14. Cells were cultured in Lonza FGM-3 Cardiac Fibroblast Growth Medium (CC-4525). To perform the CRISPR–Cas9 gene knock-out screen, on day 1 Cas9-expressing fibroblasts were seeded at 4,000 cells per well in 50 µl complete growth medium containing 10 µg ml^{-1} Polybrene (Sigma TR-1003-G) in 384-well Perkin Elmer Cell Carrier imaging microplates. This was immediately followed by transduction with 8 µl per well of lentiviral sgRNA. Negative controls included non-targeting guides and positive controls included *TGFBR1*, *TGFBR2*, *ACTA2*, *SMAD2* and *SMAD3* guides. One to four different guides were used per gene. The following day, day 2, cells were washed twice with 80 µl per well PBS and 60 µl per well selection media containing 2 µg ml^{-1} puromycin was then added to the plates. Transduction and washing of 384-well plates were accomplished using an Integra Viaflo pipetting system. Cells were incubated in selection media for 6 days to ensure complete gene knockout before performing the fibroblast activation assay. On day 8, the assay was initiated by performing a media exchange for starvation medium (CC-4525 without FBS and supplements) containing TGFβ1 (Sigma T7039) at 6 ng ml^{-1} (EC99). Cells were stimulated with TGFβ1 for 48 h to allow fibroblast–myofibroblast transition to occur.

Fixation and immunofluorescence staining of plates for imaging was carried out using a Thermo Fisher Multidrop Combi and a Biotek plate-washer. After 48 h, cells were fixed with 40 µl per well ice-cold 100% methanol for 20 min and additionally permeabilized with 40 µl per well 0.5% Triton-X 100 in PBS for 15 min. Roche Blocking Reagent at 40 µl per well was used to block cells for 15 min before addition of antibodies. The primary antibody cocktail consisted of mouse monoclonal antibody (1A4) to α-SMA (Abcam ab7817, 1:500), goat polyclonal to β-tubulin (Abcam ab21057, 1:500) and COL1A1 (E6A8E) rabbit monoclonal antibody (Cell Signaling Technology 39952S, 1:200). The primary antibody cocktail was incubated with cells at 30 µl per well for 1.5 h at room temperature. The secondary antibody cocktail consisted of donkey anti-goat IgG (H+L) highly cross-adsorbed secondary antibody, Alexa Fluor Plus 647 (Life Technologies A32849), donkey anti-mouse IgG (H+L) highly cross-adsorbed secondary antibody, Alexa Fluor 488 (Life Technologies A21202), donkey anti-rabbit IgG (H+L) highly cross-adsorbed secondary antibody, Alexa Fluor 568 (Life Technologies A10042) and Hoechst 33342 (Thermo Fisher 62249). All secondary antibodies were used at a 1:1,000 dilution and incubated with cells at 30 µl per well at room temperature for 45 min. Two washes of 80 µl per well PBS were used after each addition of fixation and permeabilization, and 4 washes of 80 µl per well PBS were used after each antibody addition.

Plates were imaged on a laser-powered Perkin Elmer Opera Phenix confocal imaging system using a water-immersion 20× objective with 9 fields of view per well. Image analysis was performed using PerkinElmer Harmony High-Content Imaging and Analysis Software. Single cells were segmented using Hoechst channel for nucleus and tubulin channel for cytoplasm. To quantify the myofibroblast phenotype, multiple intensity, STAR morphology and texture parameters for tubulin and SMA staining were measured within every cell. The PhenoLOGIC machine learning linear classifier method (Perkin Elmer) was used to differentiate between fibroblasts and activated myofibroblasts. The control plate, in which the cells were seeded with different cell numbers and treated with different concentrations of TGFβ1, was used for training. High concentration wells represented myofibroblasts and low or no TGFβ1 wells represented fibroblasts. As a readout, the fraction of myofibroblasts per well was calculated. For each guide, the mean value and standard error was calculated excluding wells with low cell counts. This low cell count cut-off per well was determined by plotting the binned cell number against the fraction of myofibroblasts per well and cell counts for which the control wells did not perform well, were accepted as not reliable. The assay was repeated up to a maximum of five times for each sgRNA in the primary screen. Only sgRNAs with more than one well of data from at least two replicates are displayed. To compare across screens, well values for the fraction of myofibroblasts within each screen were normalized to both the median of all *TGFBR1* sgRNAs ($sgRNA_{TGFBR1}$) and the median of all NTC sgRNAs ($sgRNA_{NTC}$), calculated as 100 × (fraction of myofibroblast − median($sgRNA_{NTC}$))/(median($sgRNA_{NTC}$) − median($sgRNA_{TGFBR1}$)).

The top-performing sgRNAs for *JAZF1* (BRDN0001495689), *PRELP* (BRDN0003485480), and *TGFBR1* (BRDN0000579830) were cloned into the pLentiCRISPR eSpCas9 v2 vector (SC1823, GenScript). Lentiviral particles were generated in HEK 293t/17 cells (ATCC) using Lenti-X packaging single shots (Takara) according to the manufacturer's instructions. The HEK 293T/17 cell line was obtained directly from ATCC and no mycoplasma tests were conducted. ACBRI5118 cells were transduced with the lentiviral particles in replicate, and then selected for with Puromycin the day following transduction. RNA was collected from transduced cells using the RNeasy Mini Kit (QIAgen; cat. No. 74104) with on-column DNA digestion (The RNAse-Free DNase Set; cat. no. 79254) and 500 ng of RNA were reverse transcribed into cDNA using iScript cDNA synthesis kit (BioRad; cat. No. 1708890). PrimeTime (Integrated DNA Technologies) forward and reverse primers (sequences in Supplementary Table 16) were used in quantitative PCR reaction mixtures by adding all required components according to the manufacturer's instructions (iTaq Universal SYBR Green Supermix; BioRad; 1725121) on a BioRad CFX384 Real-Time System.

Statistics and reproducibility

RNAscope and immunofluorescent micrographs are representative. Micrographs in Fig. 4b,d were selected from 9 fields of view for NTC wells ($n = 50$), *RBM20* wells ($n = 3$), *TGFBR1* wells ($n = 20$), *TGFBR2* wells ($n = 10$), *PRELP* wells ($n = 3$) and *JAZF1* wells ($n = 10$). For NTC wells, images are representative across sgRNAs that showed a high fraction of myofibroblasts. For *TGFRBR2* wells, images are representative of sgRNAs with a low fraction of myofibroblasts. Micrographs in Extended Data Figs. 8f, 9a,b and 10a are derived from one tissue section for each patient displayed. For MK167 and CD163 imaging, this was one DCM, one HCM and two NF donors (Extended Data Figs. 8f and 9a). For *DCN* and *COL22A1* imaging in primary patients, this was one DCM, one HCM and two NF donors (Extended Data Fig. 9b). For *DCN* and *COL22A1* imaging in validation patients, this was two DCM, two HCM and two NF donors (Extended Data Fig. 10).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Processed single-nucleus transcriptomic data are available through the Broad Institute's Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) under project ID SCP1303 (https://singlecell.broadinstitute.org/single_cell/study/SCP1303/). Raw sequence data are available to authorized users through dbGaP (the database of Genotypes and Phenotypes) accession number phs001539. Use of the raw sequencing data is limited to health, medical and/or biomedical purposes, including methods development research. Source data are provided with this paper.

30. Ma, Y. et al. Cardiomyocyte d-dopachrome tautomerase protects against heart failure. *JCI Insight* **4**, e128900 (2019).
31. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
32. Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: A deep generative model for unsupervised removal of background noise from scRNA-seq datasets. Preprint at *bioRxiv* https://doi.org/10.1101/791699 (2019).
33. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
34. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
35. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281–291.e9 (2019).
36. Lun, A. T. L. & Marioni, J. C. Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostatistics* **18**, 451–464 (2017).
37. Chen, Y., Lun, A. T. L. & Smyth, G. K. From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* **5**, 1438 (2016).
38. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
39. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
40. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
41. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 1–10 (2021).
42. Brill, B., Amir, A. & Heller, R. Testing for differential abundance in compositional counts data, with application to microbiome studies. Preprint at *arXiv* https://arxiv.org/abs/1904.08937 (2019).
43. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
44. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. Preprint at *bioRxiv* https://doi.org/10.1101/060012 (2016).
45. Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
46. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
47. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
48. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
49. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
50. Van den Berge, K. et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
51. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
52. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
53. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* https://arxiv.org/abs/1303.3997 (2013).
54. Poplin, R. et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at *bioRxiv* https://doi.org/10.1101/201178 (2017).
55. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
56. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **97**, 676–682 (2012).
57. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
58. Sanson, K. R. et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 1–15 (2018).
59. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

Acknowledgements The Precision Cardiology Laboratory is a joint effort between the Broad Institute and Bayer AG. This work was supported by the Fondation Leducq (14CVD01), and by grants from the National Institutes of Health to P.T.E. (1RO1HL092577, R01HL128914 and K24HL105780), N.R.T. (5K01HL140187), K.B.M. (1R01HL105993) and M.C.H. (5T32HL007208-42). This work was also supported by a grant from the American Heart Association Strategically Focused Research Networks to P.T.E. (18SFRN34110082).

Author contributions M.C., N.R.T., C.M.S. and P.T.E. conceived and designed the study. I.P., B.S., A.-D.A. and A.A. generated snRNA-seq data. M.C. and S.J.F. analysed the snRNA-seq data. I.P., M.C.H., S.H., J.Y., M.N. and P.T.E. interpreted snRNA-seq analyses. B.S. performed RNA in situ hybridization and immunofluorescence staining. M.M., M.K.-A., M.C.H., O.A. and V.K.K. developed, executed, validated and interpreted cardiac fibrosis cellular assay. K.C.B. and K.B.M. acquired LV samples for this study. M.C., I.P., B.S. and P.T.E. drafted the manuscript, which was revised by all authors.

Competing interests A.-D.A. is an employee of Bayer US LLC (a subsidiary of Bayer AG) and may own stock in Bayer AG. I.P., S.H. and C.M.S. were full-time employees of Bayer when this work was performed. C.M.S. is now a full-time employee of Vifor Pharma Management. V.K.K. was a co-founder of Dogma Therapeutics and Apolo1 Bio, and is now an employee of Anji Pharmaceuticals. P.T.E. has received sponsored research support from Bayer AG and IBM Health, and he has served on advisory boards or consulted for Bayer AG, MyoKardia and Novartis. K.B.M. has research grant funding from Sanofi-Aventis, USA and has also served on advisory boards for MyoKardia and Pfizer. M.N. has received funds or material research support from Lilly, Alnylam, Biotronik, CSL Behring, GlycoMimetics, GSK, Medtronic, Novartis and Pfizer, as well as consulting fees from Biogen, Gimv, IFM Therapeutics, Molecular Imaging, Sigilon and Verseau Therapeutics. The Broad Institute, Massachusetts General Hospital and Bayer AG have a pending patent application (US application no. 17/669,904) on the identification of activated fibroblasts for the diagnosis and treatment of cardiomyopathy based on the results from this work with M.C., I.P., A.-D.A., C.M.S. and P.T.E. listed as co-inventors. All remaining authors declare no competing interests.

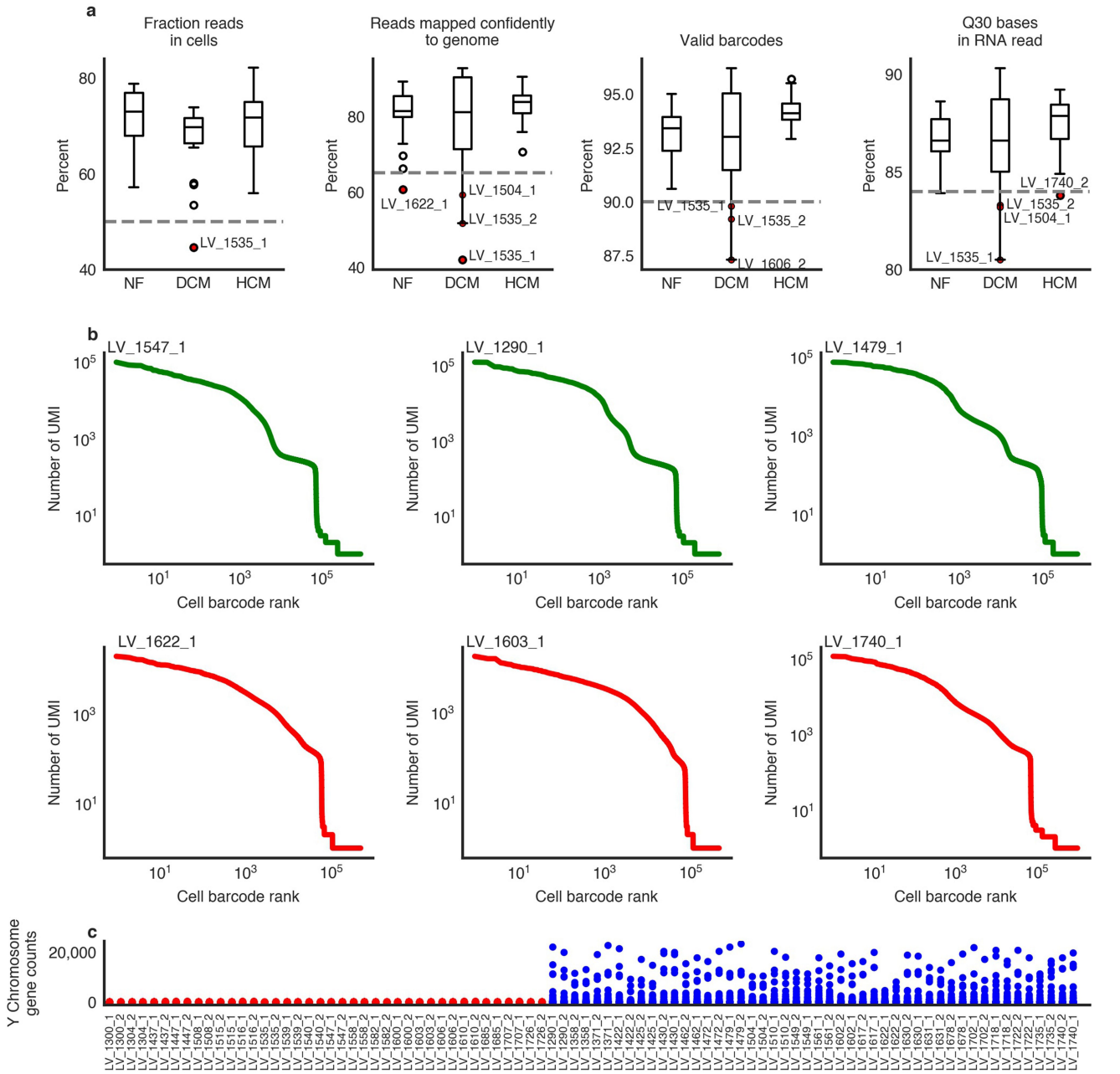
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04817-8.

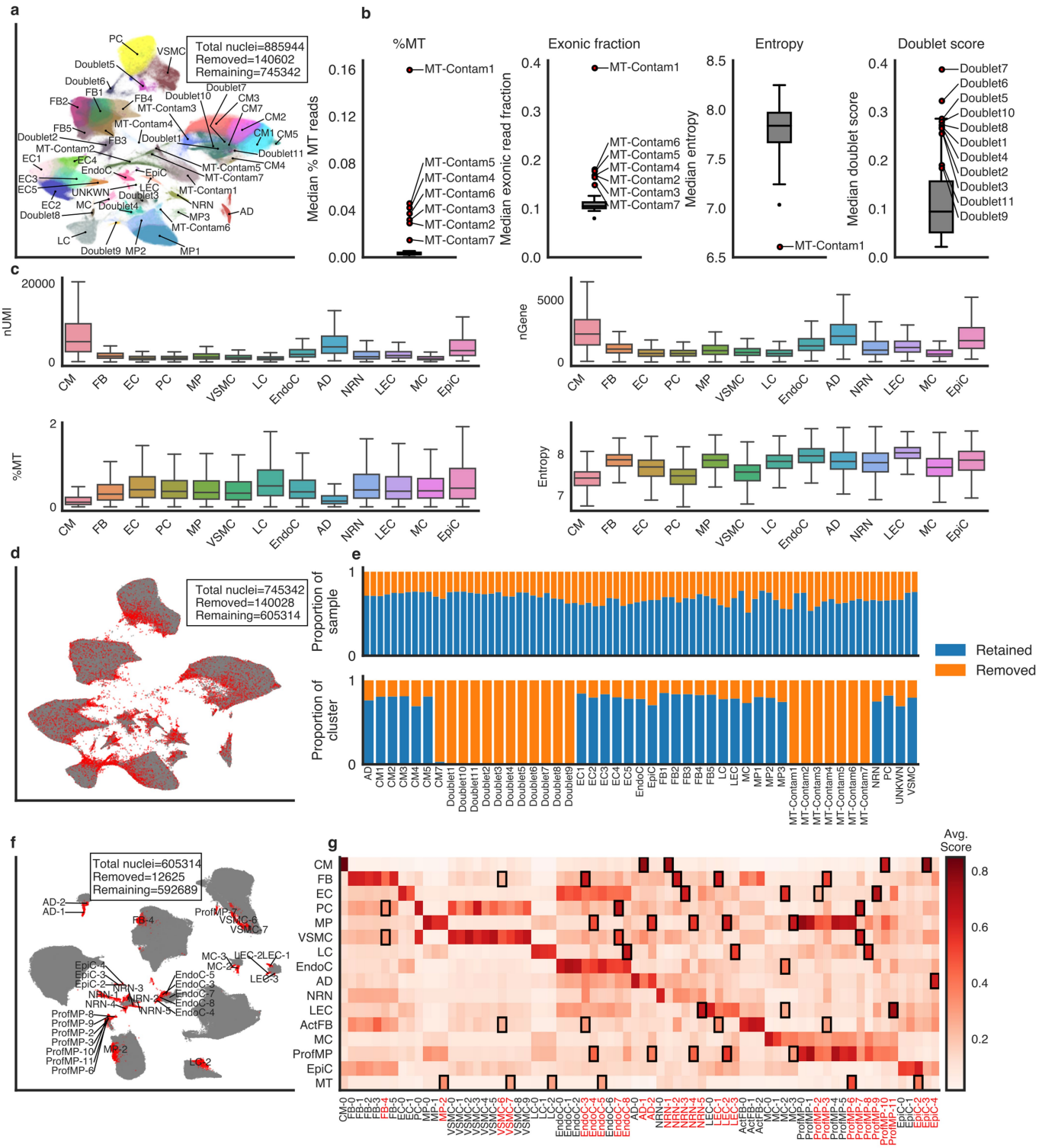
Correspondence and requests for materials should be addressed to Patrick T. Ellinor.

Peer review information *Nature* thanks Eva van Rooij and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.

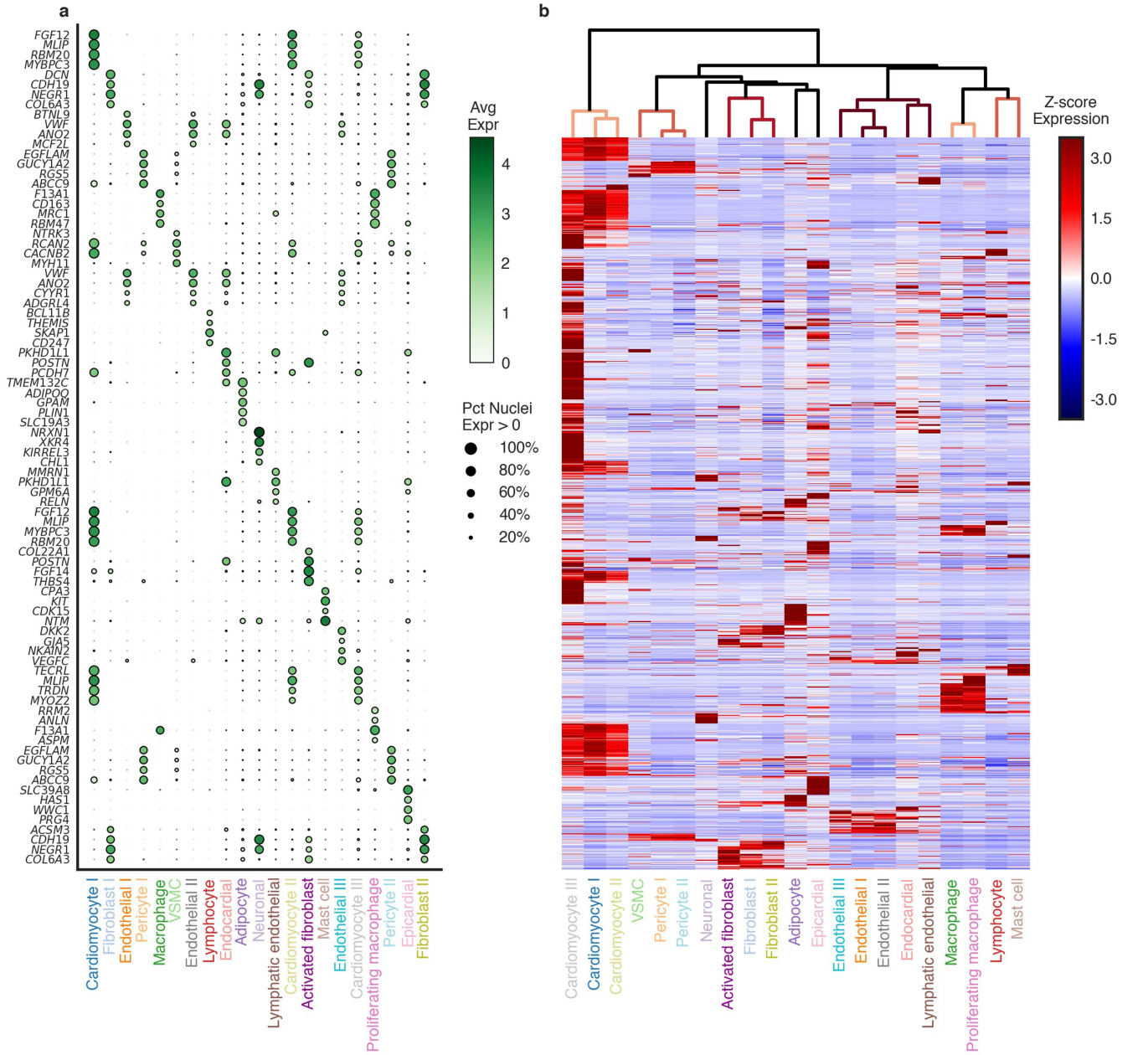


Extended Data Fig. 1 | Sample quality control assessment. a, Distribution of key quality control metrics from CellRanger count as boxplots stratified by non-failing (NF, $n = 32$), dilated cardiomyopathy (DCM, n = 24), and hypertrophic cardiomyopathy (HCM, $n = 32$). Cutoffs for low quality are shown by the dashed line, and samples that fail on a given metric are labeled. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. **b**, Exemplar high quality UMI (unique molecular identifier) decay curves (green) and low quality UMI decay curves (red) are shown. Cell barcodes are ranked by the total UMI. **c**, Total counts from Y chromosome transcripts in phenotypically classified females (red) and males (blue) across all samples ($n = 88$).

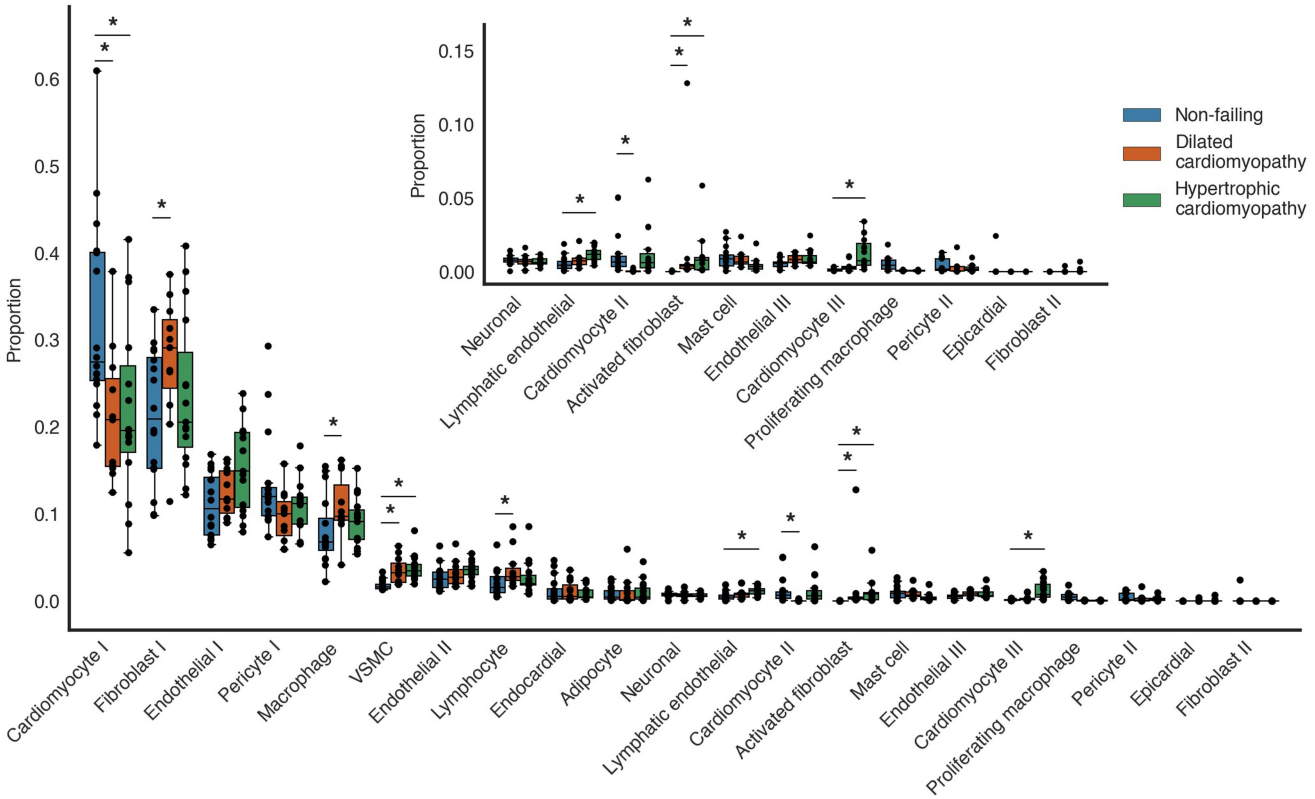


Extended Data Fig. 2 | See next page for caption.

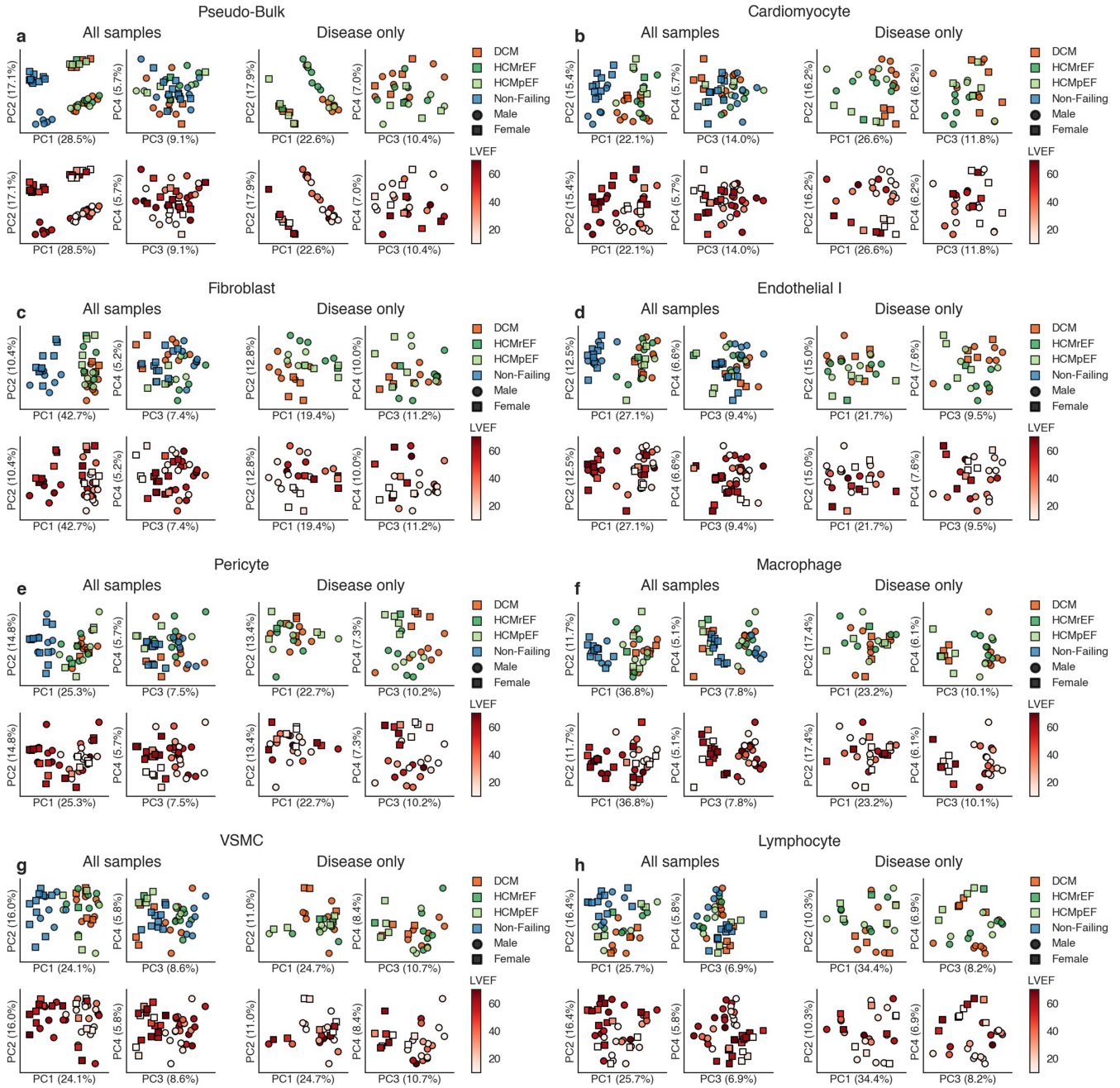
Extended Data Fig. 2 | Nuclei quality control assessment. a, Uniform manifold and projection (UMAP) representation of all CellBender non-empty droplets (n = 885,944), colored by Leiden clustering. **b**, Distribution of the median of four quality control metrics across clusters (n = 47) in **a** as boxplots, including percent of unique molecular identifiers mapping to mitochondrial genes (%MT), fraction of reads mapping exclusively to exons (Exonic fraction), entropy, and Scrublet estimated doublet score. Outlier clusters removed based on the criteria are highlighted in red. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. **c**, Distribution of number of unique molecular identifiers (nUMI), number of unique genes (nGene), %MT, and entropy across nuclei of each unique cell type (n = 13). Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. **d**, UMAP representation of non-empty droplets after removal of low-quality clusters identified in **a** and **b** (n = 745,342). Additional low-quality nuclei as detected per cluster and per-sample are colored in red. **e**, Proportion of each sample (top, n = 80) and cluster (bottom, n = 47) removed during the quality control procedure. **f**, UMAP representation of non-empty droplets after removal of low-quality clusters and per-cluster quality control (n = 605,314). Red nuclei were deemed as misclassified or low-quality nuclei based on sub-cluster analysis within each cluster. **g**, The average score for each sub-cluster (x-axis) based on marker genes for each major cell type (y-axis), see Methods. When a sub-cluster scores highly for an unrelated cell type (black border), it was removed. CM, Cardiomyocyte; FB, Fibroblast; EC, Endothelial; PC, Pericyte; MP, Macrophage; VSMC, Vascular smooth muscle cell; LC, Lymphocyte; EndoC, Endocardial; AD, Adipocyte; NRN, Neuronal; LEC, Lymphatic endothelial; ActFB, Activated fibroblast; MC, Mast Cell; ProfMP, Proliferating macrophage; EpiC, Epicardial; MT, Mitochondrial.



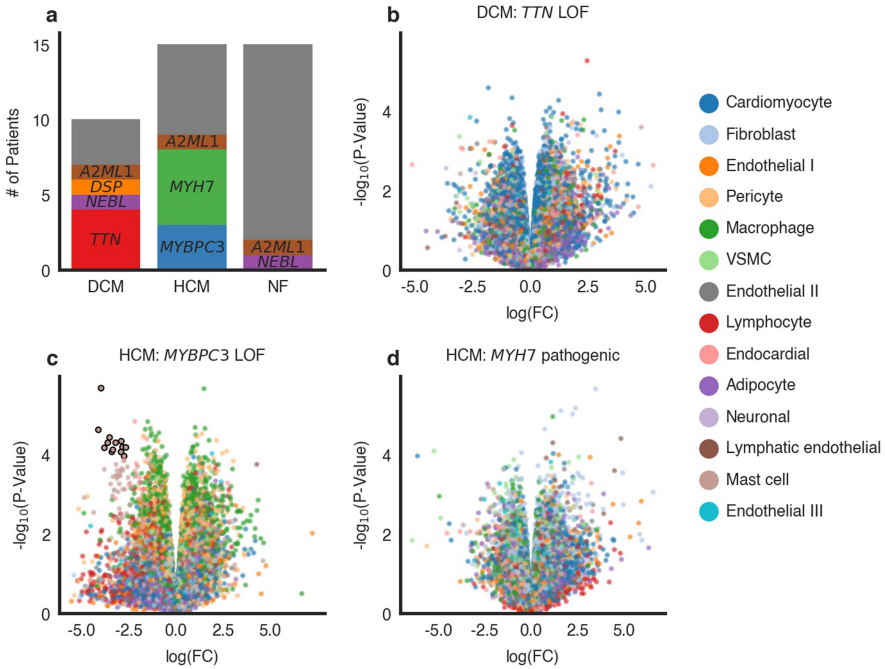
Extended Data Fig. 3 | Marker genes and cell type clustering. a, Dot plots showing the expression profile of the top 4 marker genes for each cell type in single nuclei RNA-sequencing data (n = 592,689). The size of the dot reflects the percent of nuclei expressing the gene at non-zero levels and the shade reflects the mean log-normalized expression. **b**, Hierarchical clustering of cell types and expression profiles of the top 2,000 most highly variable genes. Avg Expr, Average log-normalized expression; Pct Nuclei Expr > 0, Percent of nuclei expressing the gene at non-zero levels; VSMC, Vascular smooth muscle cell.



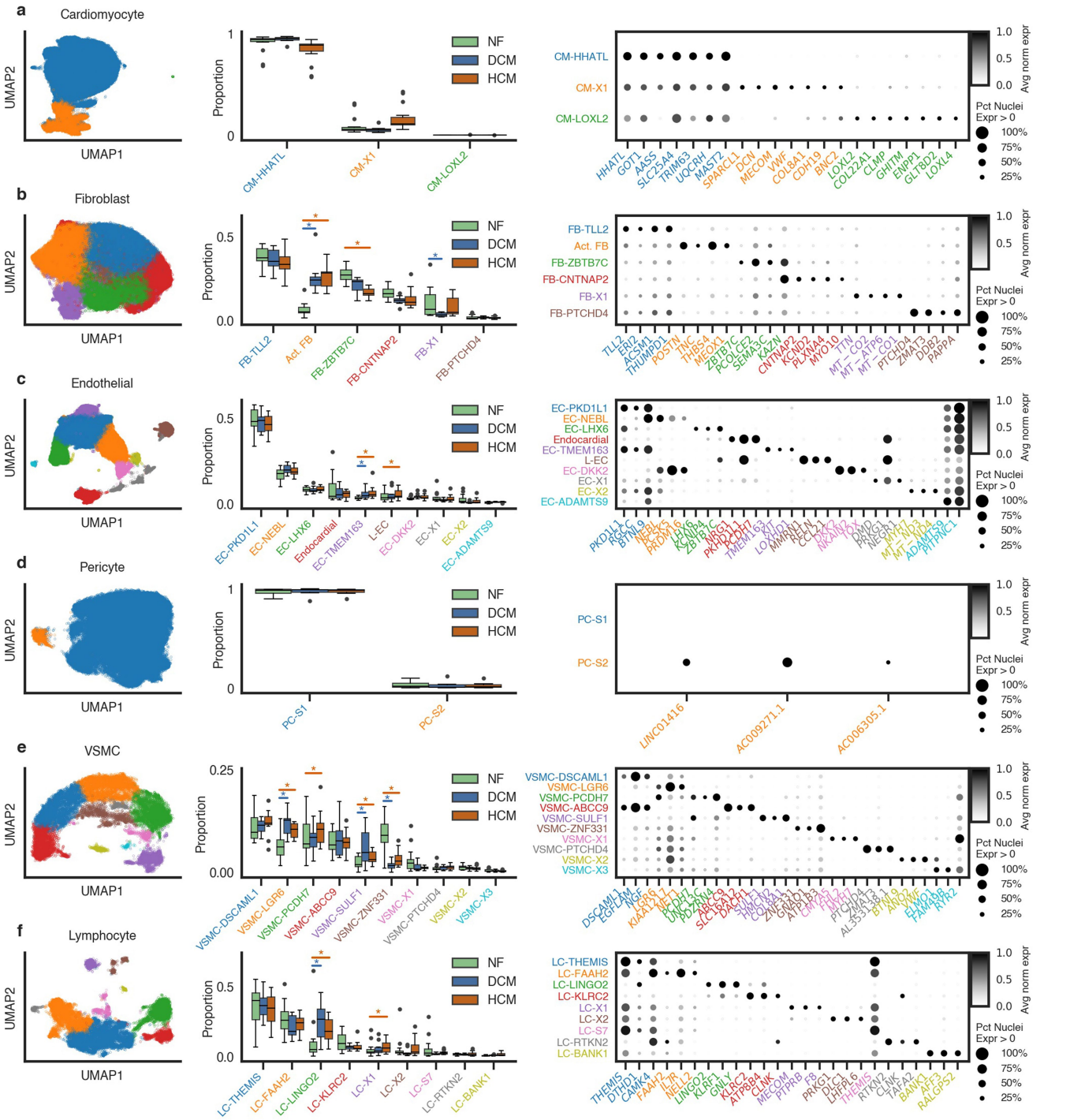
Extended Data Fig. 4 | Compositional changes of global cell types. Proportion of each cell type across patients stratified by disease status. Statistically credible changes in cell types comparing dilated cardiomyopathy ($n = 11$) or hypertrophic cardiomyopathy ($n = 15$) to non-failing patients ($n = 16$), as tested with *scCODA*, are denoted with a *, see Methods. Less prevalent cell types are shown in inset to improve readability. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; VSMC, vascular smooth muscle cell.



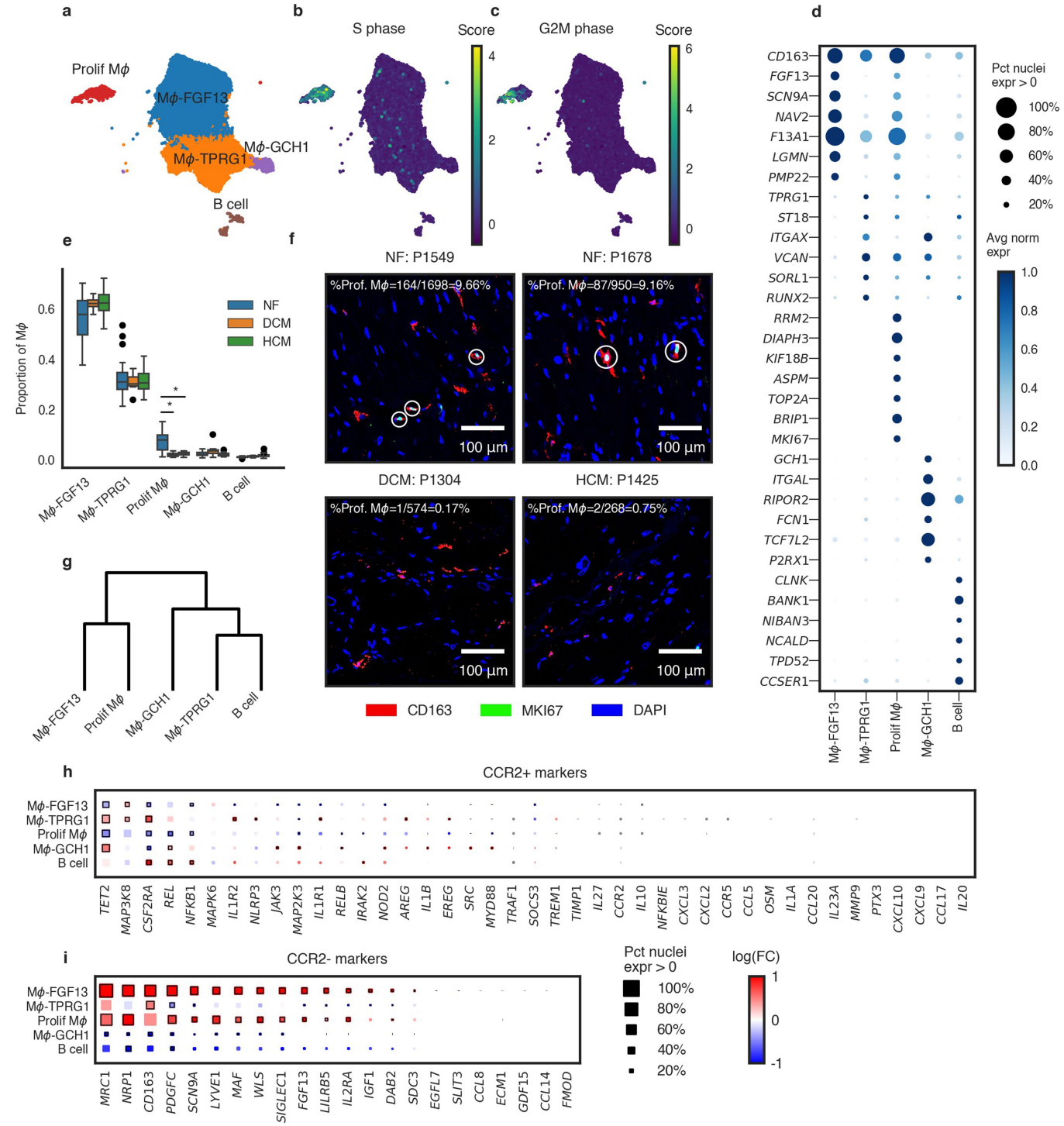
Extended Data Fig. 5 | Sample-level principal component analysis by cell type. Principal component analysis (PCA) using the top 500 most highly variables genes after summing expression counts for each sample ($n_{NF} = 16$, $n_{DCM} = 11$, $n_{HCMrEF} = 7$, $n_{HCMpEF} = 8$), for pseudo-bulk (**a**), cardiomyocyte (**b**), fibroblast (**c**), endothelial cell I (**d**), pericyte (**e**), macrophage (**f**), vascular smooth muscle cell (VSMC) (**g**), and lymphocyte (**h**). The analysis was performed across all samples (left) and restricted to cardiomyopathy samples (right). The percent of total variation for each principal component (PC) is shown in parentheses. DCM, Dilated cardiomyopathy; HCMrEF, Hypertrophic cardiomyopathy with reduced ejection fraction; HCMpEF, Hypertrophic cardiomyopathy with preserved ejection fraction; LVEF, left ventricular ejection fraction.



Extended Data Fig. 6 | Whole genome sequencing analysis of cardiomyopathy patients. a, Number of deleterious variant carriers of known clinical cardiomyopathy testing panel genes by disease state ($n_{DCM} = 10$, $n_{HCM} = 15$, $n_{NF} = 15$). Grey indicates patients with no loss-of-function (LOF) or pathogenic variant in known cardiomyopathy genes. **b**, **c**, **d** Volcano plots displaying the log fold-change (log(FC)) and two-sided *p*-value from a limma-voom differential expression analysis between dilated cardiomyopathy (DCM) *TTN* LOF carriers and DCM non-carriers ($n = 4$ vs 6) (**b**), hypertrophic cardiomyopathy (HCM) *MYBPC3* LOF carriers and HCM non-carriers ($n = 3$ vs 12) (**c**), and HCM *MYH7* pathogenic variant carriers and HCM non-carriers ($n = 5$ vs 10) (**d**) based on CellBender remove-background counts. Dots are colored by cell type with outlined dots representing genes with FDR < 0.01 as calculated with the Benjamini-Hochberg procedure. Only genes deemed to have a low probability of background contamination are displayed. VSMC, vascular smooth muscle cell.

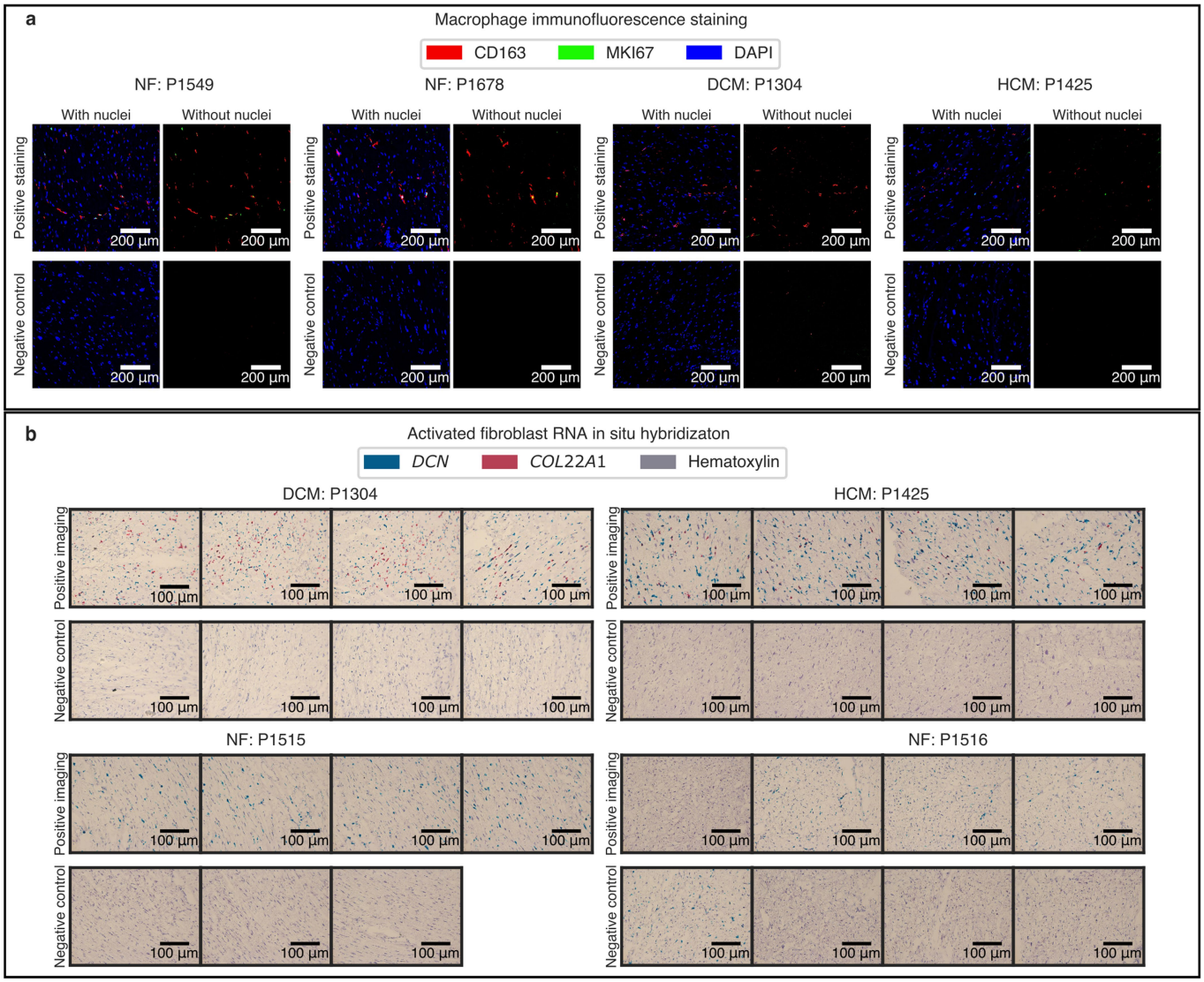


Extended Data Fig. 7 | Sub-clustering of abundant cell types. **a-f**, Sub-clustering results for cardiomyocytes (n = 158,469) (**a**), fibroblasts (n = 147,219) (**b**), endothelial cells (n = 112,977) (**c**), pericytes (n = 69,304) (**d**), vascular smooth muscle cells (VSMC, n = 18,137) (**e**), and lymphocytes (n = 16,246) (**f**). Uniform manifold approximation and projection (UMAP) visualization colored by Leiden clusters is shown on the left. The distribution of sub-populations across patients by disease status represented as box plots with statistically credible changes indicated with a * (middle, see Methods). Sub-population labels are colored as in the UMAP visualization. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. Dot plots of the most selective markers for each sub-population compared to all other sub-populations (right). The size of the dot reflects the percent of nuclei expressing the gene at non-zero levels and the shade reflects the mean log-normalized expression. Avg norm expr, Average log-normalized expression scaled to the max value for each gene; Pct Nuclei Expr > 0, Percent of nuclei expressing the gene at non-zero levels; Act. FB, Activated fibroblast; L-EC, Lymphatic endothelial.

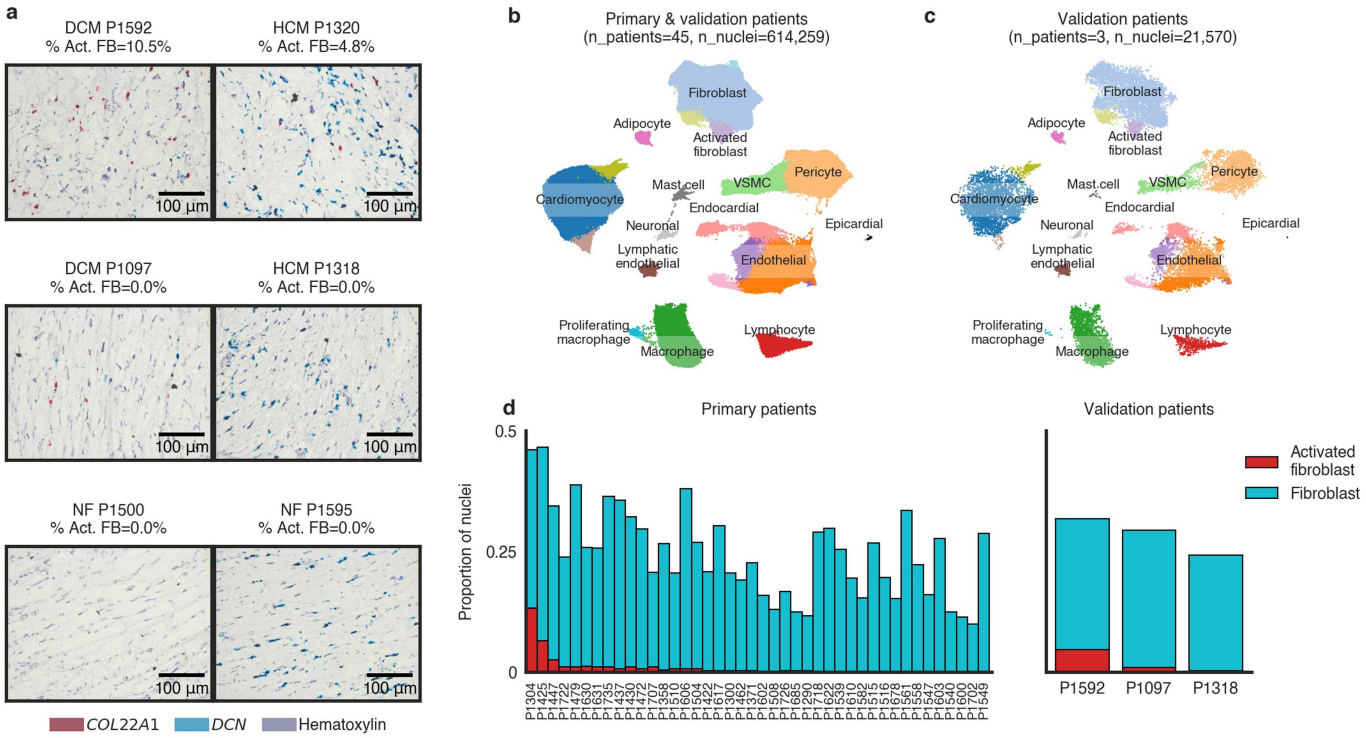


Extended Data Fig. 8 | See next page for caption.

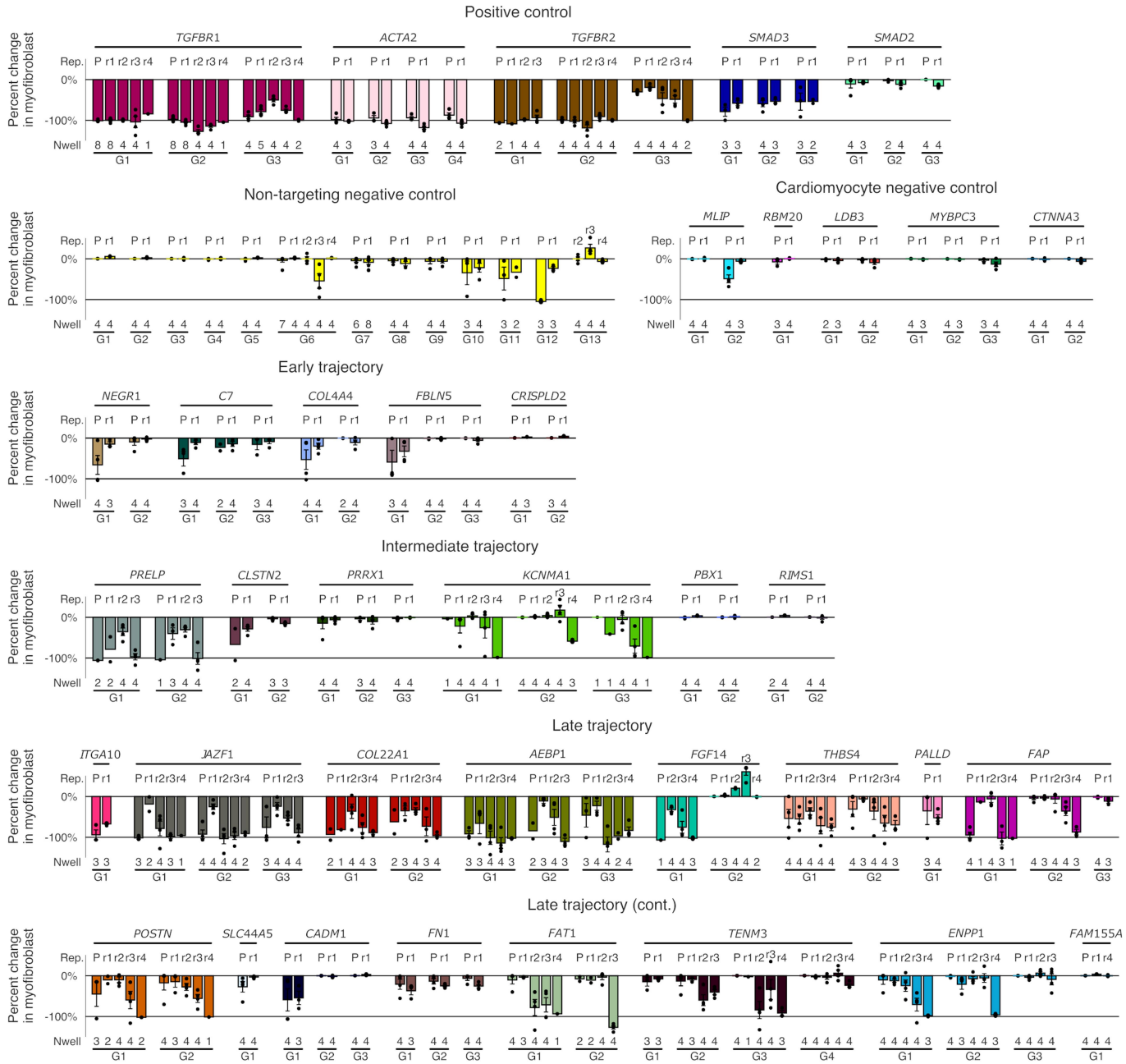
Extended Data Fig. 8 | Macrophage sub-populations. a, Uniform manifold approximation and projection (UMAP) representation of 53,730 nuclei classified as macrophage or proliferating macrophage in the global analysis colored by sub-population. **b**, UMAP plot with the S phase cell cycle score overlaid (see Methods). **c**, UMAP plot with the G2M phase cell cycle score overlaid (see Methods). **d**, Expression of markers for each sub-population. The size of each dot reflects the percent of nuclei expressing the gene at non-zero levels and the shade reflects the mean log-normalized expression. **e**, Distribution of macrophage sub-populations across patients ($n_{DCM} = 11$, $n_{HCM} = 15$, $n_{NF} = 16$) by disease status, with statistically credible differences denoted with a *, see Methods. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. **f**, Representative immunofluorescence staining for macrophage marker CD163, cycling marker MKI67, and nuclei with DAPI from single tissue sections of four patients. The percent of macrophages with MKI67 expression across full tissue sections are shown on each image. **g**, Dendrogram demonstrating similarity of sub-population centroids based on the top 2000 most highly variable genes using Euclidean distance and the Ward method. **h-i**, Expression of marker genes for CCR2 positive (CCR2+) (**h**) and CCR2 negative (CCR2-) (**i**) cardiac macrophages obtained from Bajpai et al., 2018[15]. The size of each square represents the percent of nuclei expressing the gene at non-zero levels while the shade represents a log fold-change (log(FC)) estimate comparing the expression in the given sub-population to all other sub-populations. Genes significantly up- or down-regulated in a sub-population (FDR < 0.01, as calculated with the Benjamini-Hochberg procedure) are shown with a black border. Mφ, Macrophage; NF, Non-failing; DCM, Dilated cardiomyopathy; HCM, Hypertrophic cardiomyopathy; Pct nuclei expr > 0; Avg norm expr, Average normalized expression scaled to the max expression for each gene across all sub-populations.



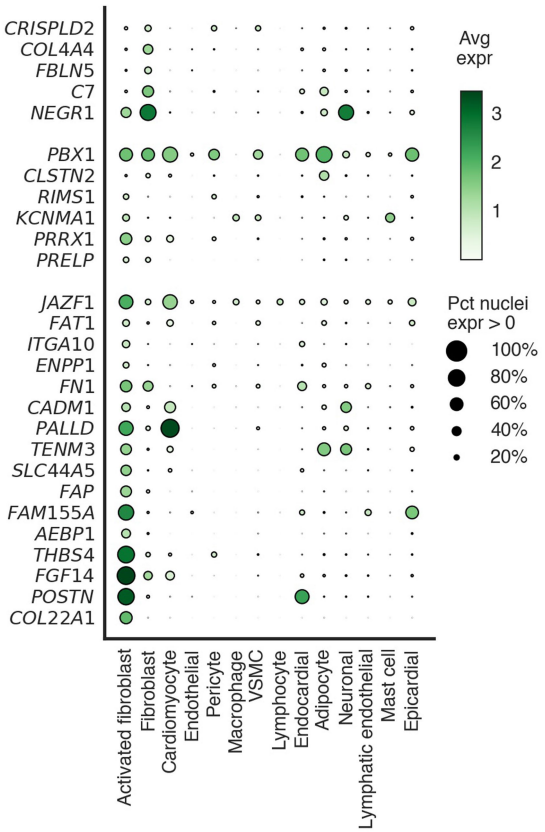
Extended Data Fig. 9 | Imaging validation. a, Additional immunofluorescence staining of four patients for macrophage marker CD163 and cycling marker MKI67 from single tissue sections of each patient shown in Extended Data Fig. 8f. Images are displayed both with, and without, DAPI to allow better visualizing of cells co-expressing CD163 and MKI67. **b**, in situ hybridization with RNAscope showing localization of canonical fibroblast marker, *DCN* (green/blue), and activated fibroblast marker, *COL22A1* (red), across several images of a single section from each of four patients: dilated cardiomyopathy patient P1304, hypertrophic cardiomyopathy patient P1425, and non-failing patients P1515 and P1516. Nuclear localization is shown with hematoxylin (blue). Negative control sections are shown in the second row for each sample. NF, non-failing; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy.



Extended Data Fig. 10 | Validation of computational deconvolution analysis for activated fibroblasts. a, *in situ* hybridization with RNAscope showing localization of canonical fibroblast marker *DCN* (green/blue), and activated fibroblast marker *COL22A1* (red), across single sections derived from 6 patients. Nuclear localization is shown with hematoxylin (blue). The computationally predicted percent of activated fibroblasts from CIBERSORTx (% Act. FB) for each patient is shown above their respective image. **b**, **c**, Uniform manifold and projection representation of all nuclei from both the primary analysis ($n_{patient}$ = 42) and validation single-nuclei RNA-sequencing ($n_{patient}$ = 3) (**b**), and separately for validation samples alone (**c**). The total number of patients (n_patients) and total nuclei (n_nuclei) are shown above each respective figure. **d**, Relative contribution of fibroblasts and activated fibroblasts to each primary analysis patient (left) and validation patients only (right). DCM, Dilated cardiomyopathy; HCM, Hypertrophic cardiomyopathy; NF, Non-failing; VSMC, Vascular smooth muscle cell.



Extended Data Fig. 11 | Replication of cardiac fibroblast activation assay across up to five screens. The percent change in the fraction of myofibroblasts across up to five screens for each sgRNA (independent sgRNAs denoted as "G"). Within each screen, well values were normalized to the median effect of *TGFBR1* sgRNAs and non-targeting control sgRNAs. The specific screen replicate (Rep.) is shown above each respective bar where "P" indicates the primary screen, and r1-r4 indicate replicate screens 1–4. The number of wells included in each screen (Nwell) are displayed below each bar with individual dots displayed for each well value. Error bars depict standard error to the mean.



Extended Data Fig. 12 | Global expression profiles of genes included in the cardiac fibroblast activation assay. Dot plot representation of expression profiles ($n_{patient}$ = 42; n_{nuclei} = 592,689) of all genes included in cardiac fibroblast activation assay. The size of the dot reflects the percent of nuclei expressing the gene at non-zero levels and the shade reflects the mean log-normalized expression. Avg expr, Average log-normalized expression; Pct nuclei expr > 0, Percent of nuclei expressing the gene at non-zero levels; VSMC, Vascular smooth muscle cell.

nature portfolio

Corresponding author(s): Patrick T. Ellinor

Last updated by author(s): Mar 21, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Harmony High-Content Imaging and Analysis Software (v4.9): https://www.perkinelmer.com/product/harmony-4-9-office-license-hh17000010
Data analysis	CellRanger (v4.0.0): https://support.10xgenomics.com/single-cell-gene-expression/software cutadapt (v1.18): https://cutadapt.readthedocs.io/en/stable/ CellBender remove-background (v0.2): https://github.com/broadinstitute/CellBender scR-Invex (sha1:4a067c5): https://github.com/broadinstitute/scrinvex Python (v3.7): https://www.python.org/ scanpy (v1.6.0): https://github.com/theislab/scanpy harmony-pytorch (v0.1.4): https://github.com/lilab-bcb/harmony-pytorch ndd (v1.6.3): https://github.com/simomarsili/ndd Scrublet (v0.2.1): https://github.com/AllonKleinLab/scrublet R (v3.5.0, v3.6.0): https://www.r-project.org/ DESeq2 (v1.20.0): https://bioconductor.org/packages/release/bioc/html/DESeq2.html limma (v3.36.5): https://bioconductor.org/packages/release/bioc/html/limma.html edgeR (v3.22.5): https://bioconductor.org/packages/release/bioc/html/edgeR.html GOstats (v2.46.0): https://bioconductor.org/packages/release/bioc/html/GOstats.html fgsea (v1.16.0): https://github.com/ctlab/fgsea ReactomePA (v1.24.0): https://bioconductor.org/packages/release/bioc/html/ReactomePA.html salmon (v1.4.0): https://github.com/COMBINE-lab/salmon scCODA (v0.1.2.post1): https://github.com/theislab/scCODA slingshot (v1.6.0): https://bioconductor.org/packages/release/bioc/html/slingshot.html Seurat (v3.1.5): https://satijalab.org/seurat/ tradeSeq (v1.2.01): https://github.com/statOmics/tradeSeq

CIBERSORTx (online interface accessed 10/28/20): https://cibersortx.stanford.edu/
velocyto (0.17.17): https://github.com/velocyto-team/velocyto.py
scVelo (v0.2.3): https://github.com/theislab/scvelo
GATK HaplotypeCaller (v3.5.0): https://gatk.broadinstitute.org/hc/en-us
BWA-MEM (v0.7.15-r1140): https://github.com/lh3/bwa
LOFTEE (v1.0): https://github.com/konradjk/loftee
hail (v0.2.62): https://github.com/hail-is/hail
Fiji is Just ImageJ (v2.1.0): https://imagej.net/software/fiji/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Processed single-nuclei transcriptomic data are available through the Broad Institute's Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) under project ID SCP1303 (https://singlecell.broadinstitute.org/single_cell/study/SCP1303/). Raw sequence data are available to authorized users through dbGaP (the database of Genotypes and Phenotypes) accession number phs001539. Publicly available bulk RNA-sequencing datasets used in this study include GEO studies GSE141910, GSE116250, and GSE130036.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We selected the largest sample set available to us of Dilated Cardiomyopathy (n=12), Hypertrophic Cardiomyopathy (n=16), and Non-failing left ventricles (n=16) for single-nuclei RNA-sequencing. No statistical methods were used to predetermine sample size.
Data exclusions	Samples were excluded if single-nuclei RNA-sequencing was low quality based on 1) <50% of reads in cells, 2) < 65% of reads confidently mapping to transcriptome, 3) < 90% valid barcodes, 4) abnormally low Q30, or 5) no ambient plateau in the unique molecular identifier (UMI) decay curve. Nuclei were removed if they 1) were enriched for mitochondrial reads, 2) were enriched for the proportion of reads mapping exclusively to exonic regions, 3) had a high prediction for being a doublet, 4) had extremely high or low UMI, 5) had extremely high or low number of genes detected, or 6) had low entropy.
Replication	We performed single-nuclei RNA-seq in duplicate on left ventricles from 44 hearts. At least 1 replicate was retained for 42 samples after quality control. Proliferating macrophage and activated fibroblast imaging validation (Extended Data Fig. 8, Extended Data Fig. 9) was performed on single sections from 1 dilated cardiomyopathy, 1 hypertrophic cardiomyopathy, and 2 non-failing patients. Consistent results were seen across cardiomyopathy patients and non-failing patients, respectively. Fibroblast trajectory analysis (Figure 3c,d,e) was replicated in two cardiomyopathy patients with consistent results. Computational deconvolution analysis (Fig. 3g, h) was performed in 3 bulk RNA-sequencing studies: GSE141910 with 360 patients, GSE116250 with 51 patients, and GSE130036 with 37 patients. Consistent enrichment was observed in cardiomyopathy patients across studies. Validation of activated fibroblast deconvolution analysis was done by performing single-nuclei RNA-sequencing and additional imaging on a subset of samples with deconvoluted bulk RNA-sequencing data (Extended Data Fig. 10). The cardiac fibroblast activation assay (Fig. 4) was performed in 5 separate experiments, with partial overlap of sgRNAs between experiments. Extended Data Fig. 11 illustrates the consistency of sgRNA effects between experiments to ensure only sgRNAs with consistent effects are discussed.
Randomization	No randomization was performed. Donors with dilated, hypertrophic cardiomyopathy, or no overt evidence of heart disease were selected for analysis. Age and sex were controlled for in differential expression analysis.
Blinding	Blinding was not relevant to our study. Given the nature of the study, we sought to compare transcriptional profiles of cardiomyopathy and non-failing left ventricles at single-nucleus resolution.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

For immunofluorescence staining validation of macrophages, commercially available primary antibodies used were: 1) anti-human Ki67 polyclonal antibody (ab15580, abcam) 1:100 dilution in 7% donkey serum overnight and 2) anti-human CD163 monoclonal antibody [OTI2G12] (ab156769, abcam) 1:100 dilution in 7% donkey serum. Secondary antibodies consisted of: 1) Alexa-fluor donkey anti-mouse 568 (A10037, Thermofisher) 1:200 dilution in 7% donkey serum and 2) Alexa-fluor donkey anti-rabbit 488 (A32790, Thermofisher) 1: 200 dilution in 7% donkey serum.

For the fibroblast cellular assay, mouse monoclonal [A14] to alpha smooth muscle actin (Abcam, ab7817/1:500), goat polyclonal to beta Tubulin (Abcam ab21057/1:500), and rabbit monoclonal COL1A1 (E6A8E) (Cell Signaling Technology 39952S/1:200) were used for primary antibody staining. The secondary antibody cocktail consisted of Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 647 (Life Technologies A32849), Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Life Technologies A21202), Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568 (Life Technologies A10042) and Hoechst 33342 (Thermo Fisher 62249). All secondary antibodies were used at a dilution of 1:1000.

Validation

All antibodies are commercially available products, with validation available at:
1. Ki67 polyclonal antibody (ab15580, abcam): https://www.abcam.com/Ki67-antibody-ab15580.html
2. CD163 monclonal antibody (ab156769, abcam): https://www.abcam.com/cd163-antibody-oti2g12-ab156769.html
3. Alexa-fluor donkey anti-mouse 568 (A10037, Thermofisher): https://www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A10037
4. Alexa-fluor donkey anti-rabbit 488 (A32790, Thermofisher): https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32790
5. Mouse monoclonal [A14] to alpha smooth muscle actin (Abcam, ab7817): https://www.abcam.com/alpha-smooth-muscle-actin-antibody-1a4-ab7817.html
6. Goat polyclonal to beta Tubulin (Abcam ab21057): https://www.abcam.com/beta-tubulin-antibody-ab21057.html
7. Rabbit monoclonal COL1A1 (E6A8E) (Cell Signaling Technology 39952S): https://www.cellsignal.com/products/primary-antibodies/col1a1-e8f4l-xp-rabbit-mab/72026
8. Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 647 (Life Technologies A32849): https://www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A32849
9. Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Life Technologies A21202): https://www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21202
10. Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568 (Life Technologies A10042): https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A10042

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

For the fibroblast cellular assay, primary human cardiac fibroblasts were used from cell line ACBRI5118 (Cell Systems): https://cell-systems.com/products/primary-human-cardiac-fibroblast-cells-acbri-5118

For lentiviral production HEK293T/17 cells were used. This cell line was acquired from the American Type Culture Collection (ATCC): https://www.atcc.org/products/crl-11268

Authentication

Human cardiac fibroblasts cells (ACBRI5118) were stained for cardiac fibroblast markers.

HEK293T/17 cells were obtained directly from ATCC.

Mycoplasma contamination

Human cardiac fibroblasts cell line (ACBRI5118) were obtained directly from Cell Systems. No Mycoplasma tests were conducted.

The HEK293T/17 cell line was obtained directly from ATCC. No mycoplasma tests were conducted.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in the study.

Human research participants

Policy information about studies involving human research participants

Population characteristics	All samples came from the Myocardial Applied Genomics Network (MAGNet) repository. Left ventricle tissue was obtained from a total of 44 individuals, including 12 dilated cardiomyopathy donors, 16 hypertrophic cardiomyopathy donors, and 16 non-failing donors. All samples were of European ancestry. 5/12 dilated cardiomyopathy donors were female, 5/16 hypertrophic cardiomyopathy donors were female, and 10/16 non-failing donors were female. Average age of dilated cardiomyopathy patients was 55, hypertrophic cardiomyopathy 49, and non-failing 57.
Recruitment	Dilated and hypertrophic cardiomyopathy left ventricle tissue was collected at time of heart transplantation. Non-failing left ventricle tissue was collected from deceased donors with no overt cardiovascular disease.
Ethics oversight	Written informed consent for research use of donated tissue was obtained from next of kin in all cases. Research use of tissues were approved by the relevant institutional review boards at the Gift-of-Life Donor Program, the University of Pennsylvania, Massachusetts General Hospital and the Broad Institute.

Note that full information on the approval of the study protocol must also be provided in the manuscript.